Shinjo Yada*

Bayesian adaptive design of early-phase clinical trials for precision medicine based on cancer biomarkers

<https://doi.org/10.1515/ijb-2021-0009>

Received November 16, 2020; accepted May 25, 2021; published online June 10, 2021

Abstract: Cancer tissue samples obtained via biopsy or surgery were examined for specific gene mutations by genetic testing to inform treatment. Precision medicine, which considers not only the cancer type and location, but also the genetic information, environment, and lifestyle of each patient, can be applied for disease prevention and treatment in individual patients. The number of patient-specific characteristics, including biomarkers, has been increasing with time; these characteristics are highly correlated with outcomes. The number of patients at the beginning of early-phase clinical trials is often limited. Moreover, it is challenging to estimate parameters of models that include baseline characteristics as covariates such as biomarkers. To overcome these issues and promote personalized medicine, we propose a dose-finding method that considers patient background characteristics, including biomarkers, using a model for phase I/II oncology trials. We built a Bayesian neural network with input variables of dose, biomarkers, and interactions between dose and biomarkers and output variables of efficacy outcomes for each patient. We trained the neural network to select the optimal dose based on all background characteristics of a patient. Simulation analysis showed that the probability of selecting the desirable dose was higher using the proposed method than that using the naïve method.

Keywords: Bayesian inference; dose-finding; phase I/II clinical trial; precision medicine.

List of abbreviations

BOIN	Bayesian optimal interval design
CR	complete response
CRM	continual reassessment method
PR	partial response
RECIST	response evaluation criteria in solid tumors

1 Introduction

Cancer is treated according to its type and location in the body, such as lung, breast, gastric, and colorectal cancers. Cancer tissue samples obtained via biopsy or surgery are normally examined for the presence of specific gene mutations by genetic testing. Next-generation sequencing allows for simultaneous analysis of

*Corresponding author: Shinjo Yada, Department of Clinical Biostatistics, Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan, E-mail: yada.shinjo.7h@kyoto-u.ac.jp

many genes at faster rates compared with conventional DNA sequencing, thus reducing the cost of genetic testing for cancer. The detected gene mutations vary based on the type of cancer, and drugs administered to treat these cancers often target specific genes. By considering not only the cancer type but also the genetic information, environment, and lifestyle of each patient, precision medicine approaches can be used to prevent and treat the disease at the individual patient level. Therefore, drug development using biomarkers for precision medicine is accelerating in oncological clinical trials [1, 2].

Several trial designs have been proposed that consider the unique characteristics of each patient [3–9]. For the trial design and method to personalize dose finding, Chen et al. [10] proposed an outcome weighted learning approach for a nonconvex loss function to estimate the optimal individualized dose (IDR). They claimed that the proposed method could be applied to infer the IDR, including accounting for drug toxicity, and that it could reveal potential biomarkers influencing the IDR. However, they did not provide specifics and concrete examples.

In clinical trials, various biomarkers are used to select a cancer treatment. As the number of patients in early clinical trials is typically limited, the accuracy of estimations of model parameters may be reduced, or it may not be possible to accurate estimations by incorporating relevant patient characteristics, including these biomarkers, in the dose-response relationship model.

To address this problem, Guo and Yuan [11] introduced an approach to reduce the number of model parameters using the canonical partial least squares (CPLS) method [12], which extracts a small number of components from the covariate matrix composed of patient characteristics. In the CPLS method, the number of components needs to be determined in advance. The number of components for the data to be analyzed must be determined using leave-one-out cross-validation or root mean squared error of prediction (RMSEP) after k-fold cross-validation. Kakurai et al. [13] proposed a dose-finding method that considers multiple genes and selects the individualized optimal dose for each patient by simultaneously performing variable selection and parameter estimation using Bayesian least absolute shrinkage and selection operator (Bayesian LASSO). They supposed that the responses were efficacy and toxicity expression. Furthermore, they supposed that dose, biomarkers, and the interaction of dose and biomarkers were explanatory variables. They fitted a Bayesian logistic regression model to estimate the efficacy probabilities for each dose and toxicity probabilities for each dose separately. Li et al. [14] proposed a utility-based approach to achieve a systematic treatment plan. For each patient, the treatment plan was specified so that the expected value of the utility function obtained by the efficacy and toxicity probabilities was maximized. They fitted a logistic regression model with dose, biomarkers, and the interaction of dose and biomarkers. They estimated efficacy and toxicity probabilities using LASSO.

In machine-learning and pattern recognition, a feature is a measurable and essential property extracted from the objected data. The CPLS method and LASSO use such features extracted by dimension reduction techniques. The approach developed by Guo and Yuan [11] and Kakurai et al. [13] applies the feature-learning method via dimension reduction in clinical trial designs. Deep learning, which has garnered attention in the field of artificial intelligence in recent years, involves learning of these features. We propose an approach for designing early-phase clinical oncology trials based on a deep neural network in deep learning model. It is expected that the generalization ability of a deep neural network model is better than that of a linear regression model, as the deep neural network model uses nonlinear activation functions. We took up a phase I/II clinical trial as an example trial design for an early-phase clinical oncology trial. We designed a Bayesian neural network model in which patient biomarkers, treatment (dose level), and interactions between dose level and biomarkers were regarded as input variables, whereas patient outcomes were regarded as target variables. The Bayesian neural network model learned from the patient data and selected the most probable optimal dose level for each patient based on the predicted results obtained using the learned network model.

2 Methodology

2.1 Network model

We established an L -layer fully connected feedforward network in which the input layer is the 0th layer and output layer is the l th layer. Let H_l denote the number of units in the l th layer, $w_{0,j}^{(l)}$ denote bias in the l th layer, and $w_{i,j}^{(l+1)}$ denote the weight parameter of the i th unit in the l th layer and j th unit in the $(l+1)$ th layer. Let $a_j^{(l+1)}$ denote an input to the j th unit in the $(l+1)$ th layer, $z_j^{(l+1)}$ denote an output from the j th unit in the $(l+1)$ th layer, and $h^{(l+1)}$ represent the activation function in the $(l+1)$ th layer ($i = 1, 2, \dots, H_l; j = 1, 2, \dots, H_{l+1}$). Suppose $x_1, x_2, \dots, x_d$ are the d input variables. We assumed the input to the j th unit in layer 1 from the input variable x_i in layer 0 (input layer) as represented by

$$a_j^{(1)} = \sum_{i=1}^d w_{i,j}^{(1)} x_i + w_{0,j}^{(1)} \quad (j = 1, 2, \dots, H_1), \quad (1)$$

and output from the j th unit of the first layer as represented by

$$z_j^{(1)} = h^{(1)}(a_j^{(1)}) \quad (j = 1, 2, \dots, H_1),$$

By adding the input data x_0 with a fixed value to 1, the bias $w_{0,j}^{(1)}$ in Eq. (1) can be included in the set of weight parameters as follows

$$a_j^{(1)} = \sum_{i=0}^D w_{i,j}^{(1)} x_i \quad (j = 1, 2, \dots, H_1).$$

Regarding the connection between the l th and $(l+1)$ th layers, we assumed the input to the j th unit of the $(l+1)$ th layer to be given by

$$a_j^{(l+1)} = \sum_{i=0}^{H_l} w_{i,j}^{(l+1)} z_i^{(l)} \quad (j = 1, 2, \dots, H_{l+1}),$$

and output from the j th unit of the $(l+1)$ th layer to be given by

$$z_j^{(l+1)} = h^{(l+1)}(a_j^{(l+1)}) \quad (j = 1, 2, \dots, H_{l+1}), \quad (2)$$

Equation (2) can be expressed using input variables $x_1, x_2, \dots, x_d$ and constant $x_0 = 1$ as

$$\begin{aligned} z_j^{(l+1)} &= h^{(l+1)}(a_j^{(l+1)}) = h^{(l+1)}\left(\sum_{i=0}^{H_l} w_{i,j}^{(l+1)} z_i^{(l)}\right) \\ &= h^{(l+1)}\left(\sum_{i=0}^{H_l} w_{i,j}^{(l+1)} h^{(l)}(a_i^{(l)})\right) \\ &= h^{(l+1)}\left(\sum_{i=0}^{H_l} w_{i,j}^{(l+1)} h^{(l)}\left(\sum_{i=0}^{H_{l-1}} w_{i,j}^{(l)} z_i^{(l-1)}\right)\right) \\ &= \dots \\ &= h^{(l+1)}\left(\sum_{i=0}^{H_l} w_{i,j}^{(l+1)} h^{(l)}\left(\sum_{i=1}^{H_{l-1}} w_{i,j}^{(l)} \dots h^{(1)}\left(\sum_{i=0}^d w_{i,j}^{(1)} x_i\right)\right)\right) \end{aligned}$$

We assumed a Bayesian neural network [15, 16], which consists of the input layer (0th layer), first hidden layer, second hidden layer, $\dots$, $(L-1)$ th hidden layer, and output layer (L th layer). Let $\mathbf{x}_n$ be the input valuable in the neural network, $y_n = y(\mathbf{x}_n, \mathbf{w})$ be the output from the neural network, and t_n be the actually observed valuable for $n = 1, 2, \dots, N$. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ represent N sets of input value in the neural network, $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$ represent N sets of output from the neural network, $\mathbf{t} = (t_1, t_2, \dots, t_N)^T$ represent

the actually observed value, and $\mathbf{w} = (w_{0,1}^{(1)}, w_{1,1}^{(1)}, \dots, w_{H_{l-1}, H_l}^{(L)})^T$ represent all parameters in the neural network. For binary classification problems, we set the input to the unit in the output layer (L th layer) as

$$a^{(L)} = \sum_{i=1}^{H_{l-1}} w_{i,j}^{(L)} z_i^{(L-1)}$$

The output from the unit in the output layer (L th layer) is

$$y_n = \sigma(a^{(L)}),$$

where $\sigma(\bullet)$ is a logistic sigmoid function. It is considered that the t_n follows the Bernoulli distribution as follows

$$p(t_n) = \sigma(a^{(L)})^{t_n} (1 - \sigma(a^{(L)}))^{1-t_n} \quad (n = 1, 2, \dots, N).$$

The likelihood function for the dataset $D = \{\mathbf{X}, \mathbf{t}\}$ is thus given by

$$p(D|\mathbf{w}, \alpha) = \prod_{n=1}^N y_n(\mathbf{x}_n, \mathbf{w})^{t_n} \{1 - y_n(\mathbf{x}_n, \mathbf{w})\}^{1-t_n}.$$

For the neural network model parameter $\mathbf{w}$, we chose a multivariate normal distribution, where the mean vector is $\mathbf{0}$ and covariance matrix is $\alpha^{-1}\mathbf{I}_W$.

$$p(\mathbf{w}|\alpha) = \frac{1}{(2\pi)^{W/2}} \frac{1}{|\alpha^{-1}\mathbf{I}_W|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{w}^T(\alpha^{-1}\mathbf{I}_W)^{-1}\mathbf{w}\right)$$

The posterior distribution of $\mathbf{w}$ $p(\mathbf{w}|D, \alpha)$ would then be given by

$$\begin{aligned} p(\mathbf{w}|D, \alpha) &= \frac{p(D|\mathbf{w}, \alpha) p(\mathbf{w}|\alpha)}{\int p(D|\mathbf{w}, \alpha) p(\mathbf{w}|\alpha) d\mathbf{w}} \propto p(D|\mathbf{w}, \alpha) p(\mathbf{w}|\alpha) \\ &= \alpha^{-W/2} \exp\left(-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\right) \prod_{n=1}^N y_n(\mathbf{x}_n, \mathbf{w})^{t_n} \{1 - y_n(\mathbf{x}_n, \mathbf{w})\}^{1-t_n} \end{aligned} \quad (3)$$

Because $y_n(\mathbf{x}_n, \mathbf{w})$ is non-linear with respect to $\mathbf{w}$, the posterior distribution could not be evaluated precisely using analytical methods. Here, we consider an approximate of (3) using the Markov chain Monte Carlo method.

2.2 Trial design

As minimal information is available for early clinical trials, determining the optimal dose level based on biomarker patterns unique to the patient at an early stage of a trial is difficult. Thus, we divided the phase I/II clinical trial into two stages, I and II, and developed a trial design that seamlessly progresses through the two stages.

The objective of the first stage is to screen for dose-based toxicity information and collect data required for application in the network model of the second stage. We used the Bayesian optimal interval design (BOIN) [17] to select the dose for the next cohort of patients based on the accumulated toxicity information, similar to the methods adopted for a conventional phase I clinical trial. Let $\pi_T^1 \leq \pi_T^2 \leq \dots \leq \pi_T^M$ be the true toxicity probability for M dose levels, $d_1 \leq d_2 \leq \dots \leq d_M$, and let ϕ_T denote the target toxicity probability. Let $\hat{\pi}_T^m$ denote the observed toxicity probability at the dose level m . $\hat{\pi}_T^m$ can be represented as $\hat{\pi}_T^m = m_m/n_m$ ($m = 1, 2, \dots, M$), with n_m and m_m denoting the number of patients dosed at level m and number of patients with observed toxicity at the dose level m , respectively. Suppose that Δ_L and Δ_U are cut-off values for the acceptable intervals of toxicity. The lower and upper bounds are given by

$$\Delta_L = \phi_T - \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right)}{\log\left\{\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right\}}, \quad \Delta_U = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right)}{\log\left\{\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right\}} - \phi_T,$$

where, ϕ_1 is the highest toxicity probability deemed as sub-therapeutic such that dose escalation is performed and ϕ_2 is the lowest toxicity probability deemed as excessively toxic such that dose de-escalation is required. The trial then flows according to the following procedure:

- (1) For the first cohort of patients, the minimum dose level is assigned.
- (2) When the dose in the current cohort is d_m :
 - (a) If $\hat{\pi}_T^m \leq \phi_T - \Delta_L$, d_{m+1} is assigned to the next cohort of patients. If d_m is d_M , d_m is assigned to the next cohort of patients.
 - (b) If $\hat{\pi}_T^m \geq \phi_T + \Delta_U$, d_{m-1} is assigned to the next cohort of patients. If d_m is d_1 , the trial is discontinued.
 - (c) If $\phi_T - \Delta_L < \hat{\pi}_T^m < \phi_T + \Delta_U$, d_m is assigned to the next cohort of patients.
- (3) We repeat (2) until drug administration to the maximum sample size of N_1 patients in stage I is complete. No patient is assigned to the dose level m that satisfies $p(\pi_T^m > \phi_T | m_m, n_m) \geq \lambda$ and $n_m \geq 3$, where λ is the cut-off to eliminate overly toxic doses. The trial is discontinued if $p(\pi_T^1 > \phi_T | m_1, n_1) \geq \lambda$ and $n_1 \geq 3$.
- (4) After drug administration to the maximum sample size of N_1 patients in stage I, from the observed data of N_1 patients using the pooling adjacent violator algorithm [18], the toxicity probability π_T^m ($m = 1, 2, \dots, M$) is estimated to satisfy the order constraint that toxicity probability increases as the dose increases. All dose levels satisfying $\hat{\pi}_T^m \leq \phi_T$ are acceptable from a safety perspective and are carried forward to stage II.

In stage II, the dose considered as optimal is assigned to each patient among the dose levels carried forward to stage II. Let x_1 denote the dose, $x_2, x_3, \dots, x_{p+1}$ denote the p biomarker status (positive or negative), y_E^m denote the actually observed binary efficacy variable (1: response, 0: non-response) at the dose level m ($m = 1, 2, \dots, M$), $\mathbf{x} = (x_1, x_2, x_3, \dots, x_{p+1}, x_1x_2, x_1x_3, \dots, x_1x_{p+1})^T$ represent the input valuables, and $y(\mathbf{x}, \mathbf{w})$ represent the output from the neural network. The efficacy probability π_E^m is expressed as:

$$\pi_E^m = p(y_E^m = 1 | \mathbf{x}, \mathbf{w}, \alpha) = \sigma(y(\mathbf{x}, \mathbf{w}))^{y_E^m} (1 - \sigma(y(\mathbf{x}, \mathbf{w})))^{1-y_E^m}$$

$$p(\mathbf{w} | \alpha) = \frac{1}{(2\pi)^{W/2}} \frac{1}{|\alpha^{-1} \mathbf{I}_W|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{w}^T (\alpha^{-1} \mathbf{I}_W)^{-1} \mathbf{w}\right)$$

where $\mathbf{w}$ is the weight parameter vector including bias in the network. For the toxicity probability π_T^m , we fit a logistic regression model with dose, biomarkers, and the interaction of dose and biomarkers as explanatory effects. We set the priors for the logistic regression model parameters to estimate the toxicity probability. The dose level m that simultaneously satisfies

$$\Pr(\pi_E^m \geq \phi_E) \geq c_f, \quad \Pr(\pi_T^m \geq \phi_T) \leq c_a \quad (4)$$

is considered as the acceptable dose from both efficacy and safety perspectives. ϕ_E is the minimum value of the expected efficacy probability, c_f is the probability cut-off corresponding to futility, and c_a is the probability cut-off corresponding to admissibility. If an acceptable dose is identified for the next patient, the dose level that maximizes the efficacy probability within the acceptable dose levels is assigned to that patient. Patients without an acceptable dose are not included in this trial. Moreover, if no acceptable dose is determined for more than a certain number of patients, the trial is discontinued.

3 Simulation studies

3.1 Simulation settings

To investigate the operating characteristics of the proposed method, simulation studies were conducted. Dose-limiting toxicity was defined based on the Common Terminology Criteria for Adverse Events (CTCAE) [19]. A binary toxicity outcome for each patient was measured, regardless of whether dose-limiting toxicities were observed. Tumor shrinkage was evaluated as a complete response (CR), partial response (PR), progressive

disease, and stable disease using the Response Evaluation Criteria in Solid Tumors (RECIST) [20]. A binary efficacy outcome for each patient was measured, regardless of whether the patient had an objective response (CR + PR). The phase I/II clinical trial was divided into two stages: stage I and stage II. In stage I, the trial was carried out based on the BOIN using patients' toxicity data. Once the pre-determined maximum sample size in stage I was reached, stage II was initiated. In stage II, the dose level that maximized efficacy among the acceptable dose levels was selected for allocation to the next patient.

The maximum sample sizes in stages I and II were set to $N_1 = 18$ and $N_2 = 27$, respectively. Five dose levels were selected in the ascending order of dosage, starting with level 1. Using three types of biomarkers (biomarkers 1, 2, and 3), positive and negative test results were obtained before selecting the dose level to be assigned to the next cohort. For each scenario, there were $2^3 = 8$ positive-negative patterns of biomarkers for each dose level.

In the proposed method, the trial proceeded with the BOIN in stage I. After drug administration to the pre-specified maximum sample size in stage I, the acceptable dose levels from a safety perspective are identified, and are carried forward to stage II. In stage II, we developed a Bayesian neural network consisting of input, hidden, and output layers. We constructed seven input units (dose level, three biomarkers, and their interactions with the dose level), two hidden units with a hyperbolic tangent activation function, and a single output whose activation function is a logistic sigmoid function. A cross-entropy error function was used with respect to a set of training data for binary classification. We used the Bayesian neural network to estimate the efficacy probabilities. The toxicity probabilities were estimated using a Bayesian logistic regression model with the dose, three biomarker status, and their interactions with the dose level. The dose level satisfying Eq. (4) was considered as the acceptable dose. We selected the dose level that maximized effective among the acceptable dose levels for allocation to the next patient by using Bayesian estimations of toxicity probabilities and efficacy probabilities. If there was an acceptable dose for the next patient, the dose level that maximized the efficacy probability within the acceptable dose levels was assigned to the next patient. Patients without an acceptable dose were not included in this trial. Moreover, if more than a certain number of patients did not have an acceptable dose, the trial was discontinued.

We compared the proposed method with the following method (named as the "naïve method"). In stage II, we used a logistic regression model with efficacy expression as the response variable and the dose, three biomarker status, and their interactions with the dose level as the explanatory variables. We used a logistic regression model with toxicity expression as the response variable and dose, biomarker status, and the interaction of dose and biomarker status as the explanatory variables. Besides these, the settings in the naïve method were the same as the proposed method.

Table 1 shows the true toxicity probabilities and efficacy probabilities under each scenario. In scenario 1, three biomarkers have main effects on efficacy, and three biomarkers interact with the dose to affect efficacy. Biomarker 3 interacts with the dose to affect toxicity. In scenario 2, three biomarkers have main effects on efficacy and three biomarkers interact with the dose to affect efficacy. Biomarkers 1 and 3 have main effects on toxicity. Biomarkers 1 and 3 interact with the dose to affect toxicity. In scenario 3, biomarker 1 has a main effect on efficacy and interacts with the dose to affect efficacy. Biomarker 2 has a main effect on efficacy, and biomarker 3 interacts with the dose to affect efficacy. Biomarkers 1 and 3 have main effects on toxicity. Biomarkers 1 and 3 interact with the dose to affect toxicity. Biomarker 2 interacts with the dose to affect toxicity. In scenario 4, biomarkers 1 and 2 have main effects on efficacy and interact with the dose to affect efficacy. Biomarker 3 has no main effect on efficacy and does not interact with the dose to affect efficacy. Three biomarkers have main effects on toxicity. Biomarker 1 interacts with the dose to affect toxicity. In scenario 5, biomarker 1 has a main effect on efficacy and interacts with the dose to affect efficacy. Biomarkers 2 and 3 have main effects on efficacy. Three biomarkers have no main effects on toxicity, and the biomarkers do not interact with the dose to affect toxicity.

We generated the outcomes of the patient with $\mathbf{X} = (X_1, X_2, X_3, X_4)$ from the following models:

$$\Pr(Y_T = 1|\mathbf{X}) = 1 - \Phi(\theta + \beta_{T,0} + \beta_{T,1}X_1 + \cdots + \beta_{T,5}X_1X_2 + \cdots + \beta_{T,7}X_1X_4)$$

Table 1: Toxicity probability and efficacy probability of five dose levels based on eight biomarker patterns under five scenarios.

Scenario	Biomarker pattern	Dose level					
		1	2	3	4	5	
1	(π_T, π_E)	(++ +)	(0.050,0.045)	(0.064,0.153)	(0.079,0.272)	(0.097,0.362)	(0.118,0.420)
		(++ -)	(0.061,0.118)	(0.108,0.303)	(0.177,0.436)	(0.268,0.507)	(0.378,0.538)
		(+ - +)	(0.050,0.095)	(0.064,0.220)	(0.079,0.328)	(0.097,0.401)	(0.118,0.446)
		(+ - -)	(0.061,0.214)	(0.108,0.395)	(0.177,0.500)	(0.268,0.548)	(0.378,0.564)
		(- + +)	(0.050,0.153)	(0.064,0.233)	(0.079,0.302)	(0.097,0.357)	(0.118,0.398)
		(- + -)	(0.061,0.306)	(0.108,0.412)	(0.177,0.472)	(0.268,0.502)	(0.378,0.515)
		(- - +)	(0.050,0.263)	(0.064,0.316)	(0.079,0.361)	(0.097,0.396)	(0.118,0.424)
		(- - -)	(0.061,0.453)	(0.108,0.511)	(0.177,0.536)	(0.268,0.543)	(0.378,0.542)
2	(π_T, π_E)	(++ +)	(0.005,0.667)	(0.005,0.743)	(0.005,0.832)	(0.006,0.919)	(0.019,0.978)
		(++ -)	(0.005,0.667)	(0.007,0.665)	(0.033,0.648)	(0.116,0.604)	(0.287,0.520)
		(+ - +)	(0.005,0.401)	(0.005,0.416)	(0.005,0.448)	(0.006,0.508)	(0.019,0.606)
		(+ - -)	(0.005,0.401)	(0.007,0.331)	(0.033,0.238)	(0.116,0.133)	(0.287,0.045)
		(- + +)	(0.005,0.441)	(0.005,0.398)	(0.005,0.336)	(0.010,0.254)	(0.023,0.158)
		(- + -)	(0.008,0.441)	(0.026,0.315)	(0.073,0.157)	(0.166,0.036)	(0.314,0.005)
		(- - +)	(0.005,0.203)	(0.005,0.131)	(0.005,0.065)	(0.010,0.021)	(0.023,0.005)
		(- - -)	(0.008,0.203)	(0.026,0.089)	(0.073,0.018)	(0.166,0.005)	(0.314,0.005)
3	(π_T, π_E)	(++ +)	(0.066,0.829)	(0.153,0.939)	(0.294,0.991)	(0.476,0.999)	(0.664,0.999)
		(++ -)	(0.031,0.774)	(0.041,0.794)	(0.053,0.812)	(0.068,0.826)	(0.086,0.833)
		(+ - +)	(0.055,0.445)	(0.098,0.586)	(0.161,0.770)	(0.247,0.930)	(0.352,0.994)
		(+ - -)	(0.025,0.368)	(0.022,0.306)	(0.020,0.230)	(0.017,0.148)	(0.015,0.073)
		(- + +)	(0.040,0.601)	(0.082,0.677)	(0.151,0.770)	(0.249,0.868)	(0.375,0.948)
		(- + -)	(0.018,0.524)	(0.018,0.395)	(0.018,0.230)	(0.018,0.080)	(0.018,0.010)
		(- - +)	(0.033,0.203)	(0.048,0.192)	(0.069,0.188)	(0.096,0.194)	(0.131,0.214)
		(- - -)	(0.014,0.152)	(0.009,0.056)	(0.005,0.009)	(0.005,0.005)	(0.005,0.005)
4	(π_T, π_E)	(++ +)	(0.005,0.345)	(0.005,0.420)	(0.005,0.464)	(0.014,0.487)	(0.037,0.499)
		(++ -)	(0.005,0.345)	(0.011,0.420)	(0.031,0.464)	(0.073,0.487)	(0.149,0.499)
		(+ - +)	(0.005,0.005)	(0.006,0.005)	(0.017,0.022)	(0.044,0.089)	(0.099,0.185)
		(+ - -)	(0.014,0.005)	(0.037,0.005)	(0.085,0.022)	(0.169,0.089)	(0.294,0.185)
		(- + +)	(0.005,0.005)	(0.005,0.005)	(0.009,0.027)	(0.020,0.095)	(0.041,0.187)
		(- + -)	(0.012,0.005)	(0.026,0.005)	(0.052,0.027)	(0.096,0.095)	(0.161,0.187)
		(- - +)	(0.006,0.005)	(0.014,0.005)	(0.031,0.005)	(0.060,0.005)	(0.108,0.037)
		(- - -)	(0.039,0.005)	(0.075,0.005)	(0.130,0.005)	(0.210,0.005)	(0.312,0.037)

Table 1: (continued)

Scenario	Biomarker pattern	Dose level				
		1	2	3	4	5
5	(π_T, π_E)	(+ + +)	(0.171, 0.288)	(0.182, 0.318)	(0.194, 0.551)	(0.206, 0.979)
		(+ + −)	(0.171, 0.318)	(0.182, 0.375)	(0.194, 0.659)	(0.206, 0.995)
		(+ − +)	(0.171, 0.228)	(0.182, 0.207)	(0.194, 0.310)	(0.206, 0.817)
		(+ − −)	(0.171, 0.254)	(0.182, 0.254)	(0.194, 0.416)	(0.206, 0.922)
		(− + +)	(0.171, 0.318)	(0.182, 0.264)	(0.194, 0.235)	(0.206, 0.294)
		(− + −)	(0.171, 0.349)	(0.182, 0.318)	(0.194, 0.330)	(0.206, 0.490)
		(− − +)	(0.171, 0.254)	(0.182, 0.165)	(0.194, 0.089)	(0.206, 0.047)
		(− − −)	(0.171, 0.282)	(0.182, 0.207)	(0.194, 0.144)	(0.206, 0.123)

Scenario 1: The efficacy probability increases with increasing dose, and the degree of rise increases with decreasing dose. The biomarker pattern has a toxicity probability that gradually rises with increasing dose. Scenario 2: A biomarker pattern in which the degree of rise in the efficacy probability increases with increasing dose, and in which the degree of rise in the efficacy probability decreases with increasing dose. A biomarker pattern in which the toxicity probability gradually rises with increasing dose. Scenario 3: A biomarker pattern in which the degree of rise in the efficacy probability increases with increasing dose, in which the degree of rise in the efficacy probability decreases with increasing dose, and in which the degree of rise in the efficacy probability does not change much even if the dose is increased. A biomarker pattern in which the toxicity probability sharply rises with increasing dose, in which the toxicity probability gradually rises with increasing dose, and in which the toxicity probability exhibits minimal changes with increasing dose. Scenario 4: The efficacy probability increases with increasing dose; however, there are biomarker patterns in which the efficacy probability at high doses falls below the minimum value for the expected efficacy probability. A biomarker pattern in which toxicity probability at high doses exceeds the target toxicity probability. Scenario 5: The efficacy probability increases with increasing dose, and the degree of rise in the efficacy probability increases as the dose increases. The toxicity probability is below the target toxicity probability at all dose levels.

$$\Pr(Y_E = 1|\mathbf{X}) = 1 - \Phi\left(\frac{\theta + \beta_{E,0} + \beta_{E,1}X_1 + \cdots + \beta_{E,5}X_1X_2 + \cdots + \beta_{E,7}X_1X_4}{\exp(\tau X_1)}\right)$$

where X_1 represents the dose, X_2, X_3 , and X_4 represent biomarkers 1–3, $\Phi(\bullet)$ is a cumulative distribution function of the standard normal distribution, and $\theta \sim N(0, \sigma^2)$ is a patient-specific random effect. The model parameters set to have true toxicity probabilities and true efficacy probabilities for each scenario are shown in Table 1. The positive occurrence rates for the three biomarkers were set to 0.50, 0.40, and 0.70. The target toxicity probability was $\phi_T = 0.30$. To apply the BOIN, the cohort size was 3 with $\phi_1 = 0.6\phi_T = 0.18$, $\phi_2 = 1.4\phi_T = 0.42$, and $\lambda = 0.95$. In this setting, (a) if the estimated toxicity probability at the dose level m is equal or less than $\phi_T - \Delta_L = 0.30 - \log((1 - 0.18)/(1 - 0.30))/\log((0.30 \times (1 - 0.18)/(0.18 \times (1 - 0.30)))) = 0.06$, the dose level $m + 1$ is assigned to the next cohort of patients; (b) if the estimated toxicity probability at the dose level m is equal or greater than $\phi_T + \Delta_U = 0.30 + \log\{(1 - 0.30)/(1 - 0.42)\}/\log[\{0.42(1 - 0.30)\}/\{0.30(1 - 0.42)\}] = 0.66$, the dose level $m - 1$ is assigned to the next cohort of patients; and (c) neither (a) or (b), the dose level m is assigned to the next cohort of patients. The minimum value for the expected efficacy probability was $\phi_E = 0.15$, probability cut-off corresponding to futility was $c_f = 0.10$, and probability cut-off corresponding to admissibility was $c_a = 0.45$. For the Bayesian neural network model, we used $N(0, 25)$ as the prior distributions for weight w . We recorded 5000 posterior samples as model parameters after burn-in of 2000 iterations. In the logistic regression model, we used independent $N(0, 25)$ as prior distribution for the model parameters. Because the posterior distributions of the model parameters cannot be analytically represented, we considered 5000 samples after 2000 burn-ins from the posterior distributions using Markov chain Monte Carlo methods to estimate the efficacy and toxicity probabilities. The simulation was carried out 1000 times for each combination of conditions.

3.2 Simulation results

Table 2 summarizes the results of the simulation studies for each biomarker pattern under various scenarios. The dose shown in bold is the optimal dose. The percentage (%) at which the optimal dose was determined after completing administration to the pre-determined sample size in stage II was shown for each dose. “None” is the proportion of trials with 10 or more off protocol cases or trials for which there were no acceptable doses.

In scenario 1, three biomarkers have main effects on efficacy and interact with dose to affect efficacy. Regardless of the statuses of the three biomarkers, the scenario is that the probability of toxicity and efficacy increases with increasing dose. In all eight patterns of biomarker status, the proposed method selected the optimal dose more frequently than the naïve method. In scenario 2, all biomarkers have main effects on efficacy and interact with the dose to affect efficacy. Unlike scenario 1, the biomarker status exhibits patterns in which the efficacy probability increases with an increasing dose and but also exhibits patterns in which the efficacy probability increases with a decreasing dose. For the biomarker pattern $(+ + -)$, the selection percentage of the optimal dose was 67.3% for the proposed method and 67.6% for the naïve method. For the biomarker pattern $(+ - -)$, the selection percentage of the optimal dose was 89.2% for the proposed method and 80.5% for the naïve method. For the biomarker pattern $(- + -)$, the selection percentage of the optimal dose was 93.2% for the proposed method and 86.8% for the naïve methods. Thus, the proposed method exhibits better performance than the naïve method. The proposed method exhibited a much higher selection percentage of the optimal dose than the naïve method in all other biomarker patterns. In scenario 3, one biomarker has a main effect on efficacy and interacts with the dose to affect efficacy. Another biomarker has a main effect on efficacy and does not interact with the dose to affect efficacy, and the other biomarker has no main effect on efficacy and interacts with the dose to affect efficacy. There were five efficacy patterns depending on the three biomarker patterns: a pattern in which the dose-response curve continues to increase with the increasing dose; a pattern in which the dose-response curve increases with the increasing dose, and plateaus at high doses; a pattern in which the dose-response curve remains virtually unchanged even when the dose is increased; a pattern in which the dose-response curve continues to decrease with increasing dose; and a

Table 2: Selection percentage of the optimal dose for each biomarker pattern under five scenarios.

Scenario	Dose level						
	1	2	3	4	5	None	
1	Biomarker pattern = (+ + +)						
	(π_T, π_E)	(0.050,0.045)	(0.064,0.153)	(0.079,0.272)	(0.097,0.362)	(0.118,0.420)	
	Proposed	0.4	0.2	0.7	1.6	96.3	0.8
	Naïve	11.7	1.2	5.4	10.8	60.2	10.7
	Biomarker pattern = (+ + -)						
	(π_T, π_E)	(0.061,0.118)	(0.108,0.303)	(0.177,0.436)	(0.268,0.507)	(0.378,0.538)	
	Proposed	0.3	0.4	9.1	77.1	12.3	0.8
	Naïve	10.2	7.3	20.9	24.4	31.1	6.1
	Biomarker pattern = (+ - +)						
	(π_T, π_E)	(0.050,0.095)	(0.064,0.220)	(0.079,0.328)	(0.097,0.401)	(0.118,0.446)	
	Proposed	0.3	1.3	1.0	1.4	95.2	0.8
	Naïve	14.8	2.3	5.3	10.0	60.5	7.1
	Biomarker pattern = (+ - -)						
	(π_T, π_E)	(0.061,0.214)	(0.108,0.395)	(0.177,0.500)	(0.268,0.548)	(0.378,0.564)	
	Proposed	0.5	0.3	14.8	76.2	7.4	0.8
	Naïve	11.3	9.1	21.0	21.8	32.9	3.9
	Biomarker pattern = (- + +)						
	(π_T, π_E)	(0.050,0.153)	(0.064,0.233)	(0.079,0.302)	(0.097,0.357)	(0.118,0.398)	
	Proposed	0.8	0.4	1.4	8.5	88.1	0.8
	Naïve	21.2	2.1	4.3	9.4	53.3	9.7
	Biomarker pattern = (- + -)						
	(π_T, π_E)	(0.061,0.306)	(0.108,0.412)	(0.177,0.472)	(0.268,0.502)	(0.378,0.515)	
	Proposed	0.5	1.9	17.0	66.1	13.7	0.8
	Naïve	21.3	8.0	18.4	17.8	29.4	5.1
	Biomarker pattern = (- - +)						
	(π_T, π_E)	(0.050,0.263)	(0.064,0.316)	(0.079,0.361)	(0.097,0.396)	(0.118,0.424)	
	Proposed	3.6	4.0	6.5	4.3	80.7	0.9
	Naïve	27.2	2.1	4.3	7.3	53.0	6.1
	Biomarker pattern = (- - -)						
	(π_T, π_E)	(0.061,0.453)	(0.108,0.511)	(0.177,0.536)	(0.268,0.543)	(0.378,0.542)	
	Proposed	2.3	1.8	17.7	74.3	3.1	0.8
	Naïve	26.1	9.6	17.0	16.5	27.2	3.6
2	Biomarker pattern = (+ + +)						
	(π_T, π_E)	(0.005,0.667)	(0.005,0.743)	(0.005,0.832)	(0.006,0.919)	(0.019,0.978)	
	Proposed	3.7	2.2	5.4	11.3	72.8	4.6
	Naïve	42.2	6.5	5.6	5.4	38.8	1.5
	Biomarker pattern = (+ + -)						
	(π_T, π_E)	(0.005,0.667)	(0.007,0.665)	(0.033,0.648)	(0.116,0.604)	(0.287,0.520)	
	Proposed	67.3	12.7	5.0	3.4	7.0	4.6
	Naïve	67.6	5.0	4.4	9.5	12.0	1.5
	Biomarker pattern = (+ - +)						
	(π_T, π_E)	(0.005,0.401)	(0.005,0.416)	(0.005,0.448)	(0.006,0.508)	(0.019,0.606)	
	Proposed	6.8	1.4	1.4	5.3	80.5	4.6
	Naïve	47.1	1.3	1.6	2.0	45.2	2.8

Table 2: (continued)

Scenario	Dose level					
	1	2	3	4	5	None
Biomarker pattern = (+ − −)						
(π_T, π_E)	(0.005, 0.401)	(0.007, 0.331)	(0.033, 0.238)	(0.116, 0.133)	(0.287, 0.045)	
Proposed	89.2	1.7	3.2	0.6	0.7	4.6
Naïve	80.5	0.1	1.4	4.6	10.0	3.4
Biomarker pattern = (− + +)						
(π_T, π_E)	(0.005, 0.441)	(0.005, 0.398)	(0.005, 0.336)	(0.010, 0.254)	(0.023, 0.158)	
Proposed	79.1	7.2	5.2	1.0	2.9	4.6
Naïve	68.3	0.8	1.0	1.1	26.4	2.4
Biomarker pattern = (− + −)						
(π_T, π_E)	(0.008, 0.441)	(0.026, 0.315)	(0.073, 0.157)	(0.166, 0.036)	(0.314, 0.005)	
Proposed	93.2	1.7	0.1	0.1	0.3	4.6
Naïve	86.8	0.2	0.7	1.3	3.9	7.1
Biomarker pattern = (− − +)						
(π_T, π_E)	(0.005, 0.203)	(0.005, 0.131)	(0.005, 0.065)	(0.010, 0.021)	(0.023, 0.005)	
Proposed	93.2	0.5	0.0	0.1	1.1	5.1
Naïve	61.4	0.1	0.0	0.3	20.1	18.1
Biomarker pattern = (− − −)						
(π_T, π_E)	(0.008, 0.203)	(0.026, 0.089)	(0.073, 0.018)	(0.166, 0.005)	(0.314, 0.005)	
Proposed	94.6	0.0	0.0	0.0	0.1	5.3
Naïve	66.2	0.0	0.0	0.0	0.3	33.5
3	Biomarker pattern = (+ + +)					
	(π_T, π_E)	(0.066, 0.829)	(0.153, 0.939)	(0.294, 0.991)	(0.476, 0.999)	(0.664, 0.999)
	Proposed	0.6	34.0	57.2	0.7	1.0
	Logistic	14.2	20.0	25.8	20.7	12.5
	Biomarker pattern = (+ + −)					
	(π_T, π_E)	(0.031, 0.774)	(0.041, 0.794)	(0.053, 0.812)	(0.068, 0.826)	(0.086, 0.833)
	Proposed	15.9	3.7	4.1	15.0	54.9
	Naïve	45.5	10.1	8.4	8.4	26.4
	Biomarker pattern = (+ − +)					
	(π_T, π_E)	(0.055, 0.445)	(0.098, 0.586)	(0.161, 0.770)	(0.247, 0.930)	(0.352, 0.994)
	Proposed	0.2	0.4	12.7	65.9	14.4
	Naïve	17.6	8.5	16.7	18.6	34.5
	Biomarker pattern = (+ − −)					
	(π_T, π_E)	(0.025, 0.368)	(0.022, 0.306)	(0.020, 0.230)	(0.017, 0.148)	(0.015, 0.073)
	Proposed	88.7	1.5	1.5	0.1	1.7
	Naïve	69.3	0.3	0.4	0.8	25.7
	Biomarker pattern = (− + +)					
	(π_T, π_E)	(0.040, 0.601)	(0.082, 0.677)	(0.151, 0.770)	(0.249, 0.868)	(0.375, 0.948)
	Proposed	0.5	1.2	8.2	65.6	18.1
	Naïve	27.0	11.1	18.1	15.4	24.2
	Biomarker pattern = (− + −)					
	(π_T, π_E)	(0.018, 0.524)	(0.018, 0.395)	(0.018, 0.230)	(0.018, 0.080)	(0.018, 0.010)
	Proposed	92.7	0.3	0.0	0.1	0.4
	Naïve	80.8	0.7	0.8	0.5	14.4

Table 2: (continued)

Scenario	Dose level					
	1	2	3	4	5	None
Biomarker pattern = (− − +)						
(π_T, π_E)	(0.033,0.203)	(0.048,0.192)	(0.069,0.188)	(0.096,0.194)	(0.131,0.214)	
Proposed	39.3	3.3	3.3	5.7	41.9	6.5
Naïve	37.7	1.8	4.1	5.3	44.5	6.6
Biomarker pattern = (− − −)						
(π_T, π_E)	(0.014,0.152)	(0.009,0.056)	(0.005,0.009)	(0.005,0.005)	(0.005,0.005)	
Proposed	78.2	0.0	0.0	0.0	0.3	21.5
Naïve	47.0	0.0	0.0	0.1	1.6	51.3
4	Biomarker pattern = (+ + +)					
	(π_T, π_E)	(0.005,0.345)	(0.005,0.420)	(0.005,0.464)	(0.014,0.487)	(0.037,0.499)
	Proposed	2.9	3.8	7.4	17.8	40.9
	Naïve	13.3	0.3	0.1	1.5	35.1
	Biomarker pattern = (+ + −)					
	(π_T, π_E)	(0.005,0.345)	(0.011,0.420)	(0.031,0.464)	(0.073,0.487)	(0.149,0.499)
	Proposed	0.2	2.1	8.7	19.2	42.6
	Naïve	12.9	0.2	1.6	5.3	31.6
	Biomarker pattern = (+ − +)					
	(π_T, π_E)	(0.005,0.005)	(0.006,0.005)	(0.017,0.022)	(0.044,0.089)	(0.099,0.185)
	Proposed	0.1	0.2	1.2	1.3	61.6
	Naïve	7.1	0.0	0.4	1.9	28.7
	Biomarker pattern = (+ − −)					
	(π_T, π_E)	(0.014,0.005)	(0.037,0.005)	(0.085,0.022)	(0.169,0.089)	(0.294,0.185)
	Proposed	0.1	0.0	0.0	9.0	45.3
	Naïve	8.4	0.2	1.9	7.6	19.4
	Biomarker pattern = (− + +)					
	(π_T, π_E)	(0.005,0.005)	(0.005,0.005)	(0.009,0.027)	(0.020,0.095)	(0.041,0.187)
	Proposed	3.7	0.8	0.8	0.5	62.9
	Naïve	11.0	0.0	0.1	1.1	26.9
	Biomarker pattern = (− + −)					
	(π_T, π_E)	(0.012,0.005)	(0.026,0.005)	(0.052,0.027)	(0.096,0.095)	(0.161,0.187)
	Proposed	0.2	0.0	0.5	0.3	70.1
	Naïve	13.7	0.1	1.0	3.2	24.8
	Biomarker pattern = (− − +)					
	(π_T, π_E)	(0.006,0.005)	(0.014,0.005)	(0.031,0.005)	(0.060,0.005)	(0.108,0.037)
	Proposed	0.1	0.0	0.0	0.0	3.8
	Naïve	3.2	0.0	0.0	0.2	8.9
	Biomarker pattern = (− − −)					
	(π_T, π_E)	(0.039,0.005)	(0.075,0.005)	(0.130,0.005)	(0.210,0.005)	(0.312,0.037)
	Proposed	0.0	0.0	0.0	0.3	1.8
	Naïve	0.4	0.0	0.1	0.9	4.2
5	Biomarker pattern = (+ + +)					
	(π_T, π_E)	(0.171,0.288)	(0.182,0.318)	(0.194,0.551)	(0.206,0.979)	(0.219,0.999)
	Proposed	0.4	0.8	1.0	4.7	85.4
	Naïve	13.0	6.8	11.6	11.4	39.0

Table 2: (continued)

Scenario	Dose level					
	1	2	3	4	5	None
Biomarker pattern = (+ + −)						
(π_T, π_E)	(0.171,0.318)	(0.182,0.375)	(0.194,0.659)	(0.206,0.995)	(0.219,0.999)	
Proposed	0.2	0.7	3.6	14.4	72.9	8.2
Naïve	11.5	7.5	12.1	11.6	40.5	16.8
Biomarker pattern = (+ − +)						
(π_T, π_E)	(0.171,0.228)	(0.182,0.207)	(0.194,0.310)	(0.206,0.817)	(0.219,0.999)	
Proposed	0.2	0.2	1.6	1.4	88.5	8.1
Naïve	13.7	5.9	8.8	10.0	41.2	20.4
Biomarker pattern = (+ − −)						
(π_T, π_E)	(0.171,0.254)	(0.182,0.254)	(0.194,0.416)	(0.206,0.922)	(0.219,0.999)	
Proposed	0.5	1.1	0.9	2.7	87.0	7.8
Naïve	14.4	6.2	10.3	10.9	42.2	16.0
Biomarker pattern = (− + +)						
(π_T, π_E)	(0.171,0.318)	(0.182,0.264)	(0.194,0.235)	(0.206,0.294)	(0.219,0.664)	
Proposed	3.9	0.4	0.6	2.9	84.2	8.0
Naïve	20.4	5.4	5.4	7.6	37.9	23.3
Biomarker pattern = (− + −)						
(π_T, π_E)	(0.171,0.349)	(0.182,0.318)	(0.194,0.330)	(0.206,0.490)	(0.219,0.914)	
Proposed	3.2	0.5	1.1	8.4	78.9	7.9
Naïve	21.5	5.8	8.3	8.5	37.5	18.4
Biomarker pattern = (− − +)						
(π_T, π_E)	(0.171,0.254)	(0.182,0.165)	(0.194,0.089)	(0.206,0.047)	(0.219,0.050)	
Proposed	72.1	1.0	1.2	0.4	17.0	8.3
Naïve	30.1	2.9	3.1	2.4	30.5	31.0
Biomarker pattern = (− − −)						
(π_T, π_E)	(0.171,0.282)	(0.182,0.207)	(0.194,0.144)	(0.206,0.123)	(0.219,0.240)	
Proposed	59.4	1.5	1.2	2.1	27.7	8.1
Naïve	33.8	4.3	2.8	4.3	32.9	21.9

Dose shown in bold is the optimal dose.

pattern in which the dose-response curve decreases with increasing dose and there is almost no response at high doses. For biomarker pattern (− − +), the selection percentage of the optimal dose was 41.9% for the proposed method and 44.5% for the naïve method. For the biomarker pattern (− + −), the selection percentage of the optimal dose was 80.8% for the naïve method and 92.7% for the proposed method, demonstrating good performance. The proposed method showed a much higher selection percentage of the optimal dose than the naïve method for the other biomarker patterns. In scenario 4, among the three biomarkers, biomarkers 1 and 2 have main effects on efficacy and interact with the dose to affect efficacy, whereas biomarker 3 has neither main effects on efficacy nor any interaction with the dose to affect efficacy. In scenario 4, if both biomarkers 1 and 2 of the eight biomarker patterns are negative, there is no acceptable dose. Even for the other biomarker patterns, there are many scenarios for which the efficacy probability slightly exceeds the expected efficacy probability even at the highest dose. When the biomarker pattern did not have an acceptable dose, both proposed method and naïve method were determined with a high probability to have “None.” For biomarker pattern (+ + +), the selection percentage of the optimal dose was 40.9% for the proposed method and 35.1% for the naïve method, and for biomarker pattern (+ + −), the selection percentage of the optimal dose was 42.6% for the proposed method and 31.6% for the naïve method. The selection percentage of the optimal dose

for the proposed method exceeded that of the naïve method. For all other biomarker patterns, the selection percentage of the optimal dose for the proposed method was much higher than that of the naïve method. Scenario 5 is a case in which biomarker 1 has a main effect on efficacy and interacts with the dose to affect efficacy, and biomarkers 2 and 3 have main effects on efficacy and do not interact with the dose to affect efficacy. For all biomarker patterns, the selection percentage of the optimal dose for the proposed method was much higher than that of the naïve method.

4 Discussion

This study was conducted to develop a dose-selection method for each patient based on the patients' biomarkers in an early clinical trial with limited number of patients. To achieve this, we built a Bayesian neural network with input variables of dose, biomarkers, and interactions between dose and biomarkers and a target variable of the outcome for each patient. We selected the most probable optimal dose for each patient based on the prediction from the neural network. It is assumed that a neural network is a multi-layer generalized linear model. To investigate performance improvement of the optimal dose selection by the neural network built, we set a logistic regression model with the dose, biomarker status, and their interactions with the dose as the explanatory variables for comparison. The simulation analysis showed that the selection percentage of the optimal dose was higher in the method using the neural network model (the proposed method) than in the method using the logistic regression model (the naïve method).

Neural networks mimic the networks of nerve cells in living organisms and are considered to originate from the perceptron proposed by Rosenblatt [21]. However, several theoretical limitations were highlighted. In the 1980s, a theory aiming to efficiently express diverse data by combining multiple features was proposed. Furthermore, the prior learning technique in deep belief networks [22] showed a major breakthrough after it was suggested that even deep networks with multiple layers (traditionally considered as difficult to learn) could be learnt effectively. Such Bayesian neural networks incorporated Bayesian concepts, where prior distributions were assigned to weight parameters in the neural network to implement learning and prediction.

Although neural networks have been applied to a wide range of fields, several problems have been observed, including (1) the tendency of overfitting when fitting a complex model with many parameters to an insufficient amount of data; (2) inability to effectively handle uncertainty when learning methods of general neural networks; and (3) difficulty in determining a policy for fitting a model to obtain data because of the large number of parameters for learning efficiency and characterizing a model. Bayesian networks can be used to address these issues regarding neural networks and are expected to reflect uncertainty and naturally limit over-fitting by assuming prior distributions to weight parameters. Furthermore, Bayesian neural networks can perform quantitative model selection by assessing marginal likelihoods.

The network expressiveness is thought to be polynomial with respect to layer width, but exponential with respect to layer depth. The local minimum of the error function is expected to decrease to the global minimum through multilayering. Therefore, establishing multiple layers is considered more efficient for enhancing network expressiveness than simply increasing the number of parameters. The results of simulation studies suggest that an optimal treatment method can be selected for each patient by applying a network that sets hidden layers with multiple units, with the patient's characteristics as the input variable and patient's outcome as the target variable, rather than the statistical model that sets the patient's characteristics as the explanatory variable and outcome as the response variable.

In simulation studies, we fixed the number of hidden layers and number of hidden units. However, for practical applications, these settings should be optimized according to the observed data. Bayesian optimization can be applied to solve this issue. Evaluating the performance of the constructed network by selecting the number of layers and number of layer units depending on various situations assumed in practice using Bayesian optimization requires further analysis.

In this study, it was assumed that the parameter that specifies the prior distributions of weight parameters (referred to as "hyperparameter" herein) is fixed and known. However, in reality, there are situations in which

it would be difficult to directly specify the hyperparameter or preferable to infer the information from the observed data. Confirmation of the operating characteristics of the proposed method when applying these methods, assuming a situation in which it is difficult to set a hyperparameter as a fixed value before starting a trial, should be further examined. In this study, the Bayesian model parameter inference procedure was used for both naïve method and proposed method when designing clinical trials. Selection of prior distributions plays an important role in the Bayesian framework [23]. Sensitive analysis should be performed before the test and the prior distributions to be used should be considered. Additionally, the dose was treated as a covariate when generating data in the simulation studies. A deep neural network model uses nonlinear activation functions for the output of the unit. Therefore, it is expected that the performance of the proposed method will be better than that of the naïve method if the dose proportionality assumption does not hold. It is necessary to evaluate the effect of a deviation from dose proportionality assumption quantitatively on operating characteristics through simulation studies.

In Bayesian original framework, the clinical trial could be terminated when it is determined that sufficient information has been obtained for decision-making [24]. In this simulation studies, the number of patients in stage I and stage II were set with reference to previous studies. For most clinical trials, the number of patients must be determined in advance. As stage I and stage II considered in this study correspond to conventional phase I and phase II clinical trials, respectively, it is conceivable to set the number of patients based on the conventional approach. It is important to understand how much the performance of the proposed method differs depending on the choice of clinical trial design parameters and prior distributions for model parameters. The effect of these for the proposed method should be subject to sensitive analysis.

In cancer clinical trials, the tumor response is often used as an endpoint for efficacy evaluation in solid tumors [25]. According to RECIST, each patient's status is classified into one of the following categories: CR, PR, progressive disease, and stable disease [18]. Dose-limiting toxicity is defined using adverse events graded from 1 to 5 according to CTCAE [19] in individual trials. In this study, efficacy data were based on whether the patient's status was classified as CR or PR to examine the usefulness of Bayesian neural networks over an early-phase clinical oncology trial for simplicity. As an extension of the proposed method, efficacy and toxicity data could be considered as neural network targets, and dose allocation to patients could be performed by adaptive randomization among acceptable doses. It will be possible to evaluate the performance by comparing the extended version of the proposed method with the method in previous research. In addition, to leverage patients' data to take advantage of the information, both efficacy and toxicity data may be regarded as ordinal scaled data and a model may be presented that considers the relationship between efficacy and toxicity. A Bayesian neural network may be applied that assumes a bivariate normal distribution of ordered categorical data. Guo and Yuan [11] provided a method for a phase I/II clinical trial with efficacy and toxicity regarded as ordinal variables. By expanding the proposed method to handle such cases, it will be possible to evaluate its performance in comparison with the method by Guo and Yuan [11].

The main method for dose selection considering patient biomarkers are to (i) select the necessary variables from the explanatory variables (variable selection), and (ii) convert to lower dimensional data while maintaining the original information (dimension reduction). The studies applying these methods to clinical trial designs have been published [11, 13, 14]. It is an important work to be tackled in the future to investigate the operation characteristics by comparing the method proposed in the previous research with the method using a Bayesian neural network.

The basis for incorporating novel biomarkers into a clinical trial is most often the findings of previous studies. Multiple biomarkers are used because cancer does not develop from a single cause and its pathophysiology varies [26].

By verifying the sensitivity and specificity of biomarkers objectively, it is possible to clarify the test and diagnostic abilities of biomarkers. In simulation studies, three biomarkers were set with reference to previous studies [27–29]. In practice, it is necessary to select the biomarkers to be included in the clinical trial, focusing on the biomarkers approved or recommended by the regulatory guidelines, and to evaluate the performance of the proposed method.

Development of useful biomarkers is the key to successful personalized medicine in cancer treatment [30], and the number of clinical trials related to personalized medicine is increasing. Administered drugs expected to exert a therapeutic effect by acting on a gene mutation associated with a disease have been investigated in clinical trials. However, the potential sample size for early clinical trials is limited, and it remains difficult to use statistical models to support dose selection considering patient biomarkers. In this study, we proposed a Bayesian neural network model to overcome these problems. The results of the simulation studies suggest that the proposed method has a higher probability of selecting a dose combination with a higher therapeutic effect than the naïve method using the conventional logistic regression model. The application of neural networks to clinical trial design may be useful when administering personalized medicine for cancer treatment.

Author contribution: The author has accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The author declares no conflicts of interest regarding this article.

References

- Jardim DLF, Schwaederle M, Wei C, Lee JJ, Hong DS, Eggermont AM, et al. Impact of a biomarker-based strategy on oncology drug development: a meta-analysis of clinical trials leading to FDA approval. *J Natl Cancer Inst* 2015;107:djv253.
- Milne CP, Cohen JP, Chakravarthy R. Market watch: where is personalized medicine in industry heading? *Nat Rev Drug Discov* 2015;14:812–3.
- O’Quigley J, Shen LZ, Gamst A. Two-sample continual reassessment method. *J Biopharm Stat* 1999;9:17–44.
- O’Quigley J, Paoletti X. Continual reassessment method for ordered groups. *Biometrics* 2003;59:430–40.
- Yuan Z, Chappell R. Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clin Trials* 2004;1:499–508.
- Ivanova A, Wang K. Bivariate isotonic design for dose-finding with ordered groups. *Stat Med* 2006;25:2018–26.
- Babb JS, Rogatko A. Patient specific dosing in a cancer phase I clinical trial. *Stat Med* 2001;20:2079–90.
- Whitehead J, Zhou Y, Hampson L, Ledent E, Pereira A. A Bayesian approach for dose-escalation in a phase I clinical trial incorporating pharmacodynamics endpoints. *J Biopharm Stat* 2007;17:1117–29.
- Thall PF, Nguyen HQ, Estey EH. Patient-specific dose finding based on bivariate outcomes and covariates. *Biometrics* 2008;64:1126–36.
- Chen G, Zeng D, Kosorok MR. Personalized dose finding using outcome weighted learning. *J Am Stat Assoc* 2016;111:1509–21.
- Guo B, Yuan Y. Bayesian phase I/II biomarker-based dose finding for precision medicine with molecularly targeted agents. *J Am Stat Assoc* 2017;112:508–20.
- Indahl U, Liland K, Nas T. Canonical partial least squares — a unified PLS approach to classification and regression problems. *J Chemom* 2009;23:495–504.
- Kakurai Y, Kaneko S, Hamada C, Hirakawa A. Dose individualization and variable selection by using the Bayesian lasso in early phase dose finding trials. *J R Stat Soc C* 2019;68:445–60.
- Li P, Taylor JMG, Kong S, Jolly S, Schipper MJ. A utility approach to individualized optimal dose selection using biomarkers. *Biom J* 2020;62:386–97.
- Denker J, Schwartz D, Wittner B, Solla S, Howard R, Jackel L, et al. Large automatic learning, rule extraction, and generalization. *Complex Syst* 1987;1:877–922.
- Tishby N, Levin E, Solla S. Consistent inference of probabilities in layered networks: predictions and generalization international joint conference on. *Neural Network* 1989;2:403–9.
- Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *J Roy Stat Soc C* 2015;64:507–23.
- Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD. *Statistical inference under order restrictions*. London, New York: John Wiley & Sons; 1972.
- National Cancer Institute. Common terminology criteria for adverse events v5.0 (CTCAE); 2017. Available from: https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc_50.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumors: revised RECIST guideline. 1.1 version. *Eur J Canc* 2009;45:228–47.
- Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958;65:386–408.
- Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput* 2006;18:1527–54.

23. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
24. Berry DA. A case for Bayesianism in clinical trials. *Stat Med* 1993;12:1377–93.
25. Green S, Smith A, Benedetti J, Crowley J. Clinical trials in oncology, 3rd ed. New York: Chapman & Hall/CRC Press; 2012.
26. Martin SK, Vaughan TB, Atkinson T, Zhu H, Kyprianou N. Emerging biomarkers of prostate cancer (Review). *Oncol Rep* 2012;28:409–17.
27. Davis JE, Nemesure B, Mehmood S, Nayi V, Burke S, Brzostek SR, et al. Her2 and Ki67 biomarkers predict recurrence of ductal carcinoma in situ. *Appl Immunohistochem Mol Morphol* 2016;24:20–5.
28. Severi G, FitzGerald LM, Muller DC, Pedersen J, Longano A, Southey MC, et al. A three-protein biomarker panel assessed in diagnostic tissue predicts death from prostate cancer for men with localized disease. *Cancer Med* 2014;3:1266–74.
29. Fung KY, Tabor B, Buckley MJ, Priebe IK, Purins L, Pompeia C, et al. Blood-based protein biomarker panel for the detection of colorectal cancer. *PloS One* 2015;10:e0120425.
30. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med* 2017;377:62–70.