



A Curve-Free Bayesian Decision-Theoretic Design for Phase Ia/Ib Trials Considering Both Safety and Efficacy Outcomes

Shenghua Fan¹ · Bee Leng Lee² · Ying Lu³

Received: 13 September 2019 / Revised: 10 February 2020 / Accepted: 6 March 2020 /
Published online: 26 March 2020
© International Chinese Statistical Association 2020

Abstract

A curve-free, Bayesian decision-theoretic two-stage design is proposed to select biological efficacious doses (BEDs) for phase Ia/Ib trials in which both toxicity and efficacy signals are observed. No parametric models are assumed to govern the dose–toxicity, dose–efficacy, and toxicity–efficacy relationships. We assume that the dose–toxicity curve is monotonic non-decreasing and the dose–efficacy curve is unimodal. In the phase Ia stage, a Bayesian model on the toxicity rates is used to locate the maximum tolerated dose. In the phase Ib stage, we model the dose–efficacy curve using a step function while continuing to monitor the toxicity rates. Furthermore, a measure of the goodness of fit of a candidate step function is proposed, and the interval of BEDs associated with the best fitting step function is recommended. At the end of phase Ib, if some doses are recommended as BEDs, a cohort of confirmation is recruited and assigned at these doses to improve the precision of estimates at these doses. Extensive simulation studies show that the proposed design has desirable operating characteristics across different shapes of the underlying true toxicity and efficacy curves.

Keywords Bayesian adaptive design · Biological efficacious doses · Efficacy signals · Phase I trials · Toxicity outcomes

✉ Ying Lu
ylu1@stanford.edu

¹ Department of Statistics and Biostatistics, California State University, East Bay, Hayward, CA 94542, USA

² Department of Mathematics and Statistics, San Jose State University, San Jose, CA 95192, USA

³ Department of Biomedical Data Science, School of Medicine, Center for Innovative Study Designs and the Biostatistics Core, Stanford Cancer Institute, Stanford University, Stanford, CA 94305, USA

1 Introduction

As the first application of a new drug or drug combination to humans, a phase I oncology trial traditionally focuses on finding the safe dose or schedule of the medication. Such a trial usually is a single-arm, open label, sequential study involving a small number of patients. A principal goal of a phase I trial is to establish the recommended dose and/or schedule of an experimental drug for efficacy testing later. It is believed that the toxicity and efficacy of cytotoxic chemotherapies increase with dose; thus, the recommended dose is often selected as the maximum tolerated dose (MTD). A comprehensive review of the dose escalation methods for MTD in phase I cancer clinical trials can be found in Tourneau et al. [17].

Emerging targeted therapies, including monoclonal antibodies and immunotherapies, have distinct response outcomes from the chemotherapies. While it may be reasonable to assume monotonic toxicity with dose, the dose-limiting toxicity (DLT) may not exhibit at the therapeutic doses. Moreover, the efficacy may not follow monotonic patterns [6]. Therefore, methods that are optimal for the selection of MTD may not be appropriate for the determination of a recommended dose for targeted therapies [3]. These new therapies need alternative or complementary strategies to help the identification of dose ranges based on effect biomarkers, which could be target engagement, pharmacodynamics or disease progression biomarkers. The dose that demonstrates a certain level of biological effect measured by such a biomarker is considered a biological efficacious dose (BED), which should have a desirable efficacy performance while still safeguarding patients with an acceptable toxicity profile [14]. In a review of phase I trials that lead to a recommended phase II dose for molecularly targeted agents, Hansen et al. [7] observed that only 13% (22/161) of these trials use non-toxicity endpoints and that BEDs of molecularly targeted agents may not reach MTD in these phase I trials. Therefore, they replaced the DLT rate for a 3 + 3 design with an efficacy response rate and searched for a BED by modifying the 3 + 3 rule. Another commonly used dose target is the biological optimal dose (BOD), the optimal BED that has the highest efficacy among doses below the MTD [1].

To look for BOD, Thall and Cook [16] introduced EffTox, a Bayesian adaptive dose-finding design that models correlated binary efficacy and toxicity outcomes and incorporates the trade-off between efficacy and toxicity. Brock et al. [2] searched the PubMed on October 2016 for articles that cited Thall and Cook [16] and found that, among the 54 articles returned, only three used the EffTox design in clinical trials. This reflects the statement made by Thall [15]: “Bayesian models for early phase clinical trials have seen limited use in clinical practice”. We believe that, in general, parametric models and computational complexities are the primary barriers to widespread adoption of Bayesian designs.

Many researchers have started to develop and promote Bayesian designs that are relatively easy to understand and implement. One way is to adopt curve-free models for dose–response relations. For example, to identify the MTD, Fan et al. [5] proposed a Bayesian design that assumes only a monotonic relationship between dose and toxicity. However, for trials with both toxicity and efficacy

outcomes, there is no simple curve-free assumption on the association between toxicity and efficacy. Therefore, instead of modelling the toxicity–efficacy relation, Ivanova [8], Fan and Chaloner [4], and Zhang et al. [19] modelled toxicity and efficacy as a single trinary response variable with three outcomes: no toxicity and no efficacy (no response), efficacy and no toxicity (success), and toxicity. Fan and Chaloner [4] adopted the continuation-ratio model and provided the analytic forms of its Bayesian optimal designs. However, their target dose is the most successful dose, different to BOD. To identify BOD, Sachs et al. [14] advocated the idea of a two-stage approach, in which stage 1 (Phase Ia) focuses primarily on toxicity to assess safety and tolerability as well as to establish the MTD, and stage 2 (Phase Ib) focuses primarily on efficacy to find a BED that is lower than the MTD. In this way, the complexity of joint modelling of the toxicity–efficacy association can be avoided by utilizing the marginal distributions of toxicity and efficacy sequentially. Zang and Lee [18] adopted the idea to identify the BOD, and their design assumes only that the (marginal) dose–toxicity relationship is monotonic and the (marginal) dose–efficacy relationship is unimodal or plateaued. A common thread that runs through these works is that the designs are curve free, that is, the dose–toxicity, dose–efficacy, and toxicity–efficacy relationships are not assumed to follow specific parametric curves.

While numerous trial designs for identifying the MTD or BOD have been proposed in the literature, designs for identifying the BEDs are limited. In this paper, we propose a two-stage, curve-free Bayesian design to identify the set of all BEDs in the dose domain of the study. Our two-stage approach is similar to Zang and Lee [18] but the goal is different: instead of finding the BOD, which is the optimal BED, we seek to identify all the BEDs. The reason for this goal is that the recommended phase II dose should be a decision that involves not only the statisticians but also the medical experts (such as the clinicians and pharmacologists). By providing a set of BEDs, the medical experts have the option of selecting a BED based on pharmacology and other clinical knowledge and experience to move forward to phase II trials, rather than relying solely on statistical criteria.

The disadvantage of a two-stage approach is that without utilizing the association between the toxicity and efficacy outcomes, a larger sample size may be required. However, this disadvantage can be minimized by using an efficient algorithm in the first stage. The algorithm proposed by Fan et al. [5] has been demonstrated to locate the MTD rapidly and robustly and so we adopt it in the first stage of our design. As for the second stage, because the goal is to identify all BEDs rather than the BOD, the grouping of doses, or the pooling of the samples at those doses, to estimate a single pooled efficacy rate should lead to an increased precision of the estimate. Based on this consideration, the efficacy curve is assumed to be unimodal or plateaued, and it is modelled nonparametrically by a step function. The dose selection in the second stage is based on the estimated step function and the posterior expected utilities.

Another practical issue is the selection of the single BOD from the multiple BEDs for phase II trials. Although a phase Ia/Ib trial does not aim to make statistical inferences on the efficacy of the selected dose(s), relatively large sample sizes on the selected doses (BEDs) could be required or essential for clinicians to make a good decision. To achieve that, stopping rules are implemented and the saved sample

size can be used for a confirmation cohort. In our simulation studies, confirmation cohorts are guaranteed by lowering the maximal sample size of the first two stages.

Finally, the simplicity of the prior distributions and an intuitive way to specify the prior parameters are important for the adoption of a Bayesian method in actual trials. We choose the flexible beta priors for the toxicity/efficacy rates, which can be easily specified based on a preliminary guess of the toxicity/efficacy rates. The curve-free toxicity/efficacy models and less-informative priors can increase the robustness of the proposed design, so that errors in the initial prior specification can be quickly adjusted by the data. To facilitate the application of the proposed design, an R package has been developed and is available upon request.

The rest of our paper is organized as follows. Section 2 provides a description of the design, including the modelling of the dose–toxicity and dose–efficacy relationships, and dose selection rules based on Bayesian decision theory. In Sect. 3, we evaluate the operating characteristics of our design through simulation studies and sensitivity analysis. To promote the application of our design, a completed trial is simulated with complete details and presented in Sect. 4. Finally, we present a discussion in Sect. 5.

2 The General Design

We first describe the idea of the proposed design. It starts with identifying the safe doses. The Bayesian algorithm originally proposed by Fan et al. [5] for dose-finding or phase Ia trials, hereafter referred to as the FLW algorithm, is used to locate the MTD. Due to page limits, we cannot describe the FLW algorithm in detail here. In essence, the FLW algorithm assumes only a monotonic dose–toxicity relationship and estimates the MTD with an isotonic procedure that uses the observed as well as extrapolated toxicity data. Furthermore, it recommends the dose with the maximum posterior probability of being the MTD. Simulation studies provided by Fan et al. [5] show that the algorithm selects the target dose more frequently with fewer number of patients than other commonly used Bayesian methods; moreover, it is robust against prior misspecification. An R package that implements the FLW algorithm is available from the authors upon request.

Once the MTD has been identified with sufficient confidence, in terms of the posterior probability of the recommended dose being the MTD exceeding a pre-set threshold, the phase Ib stage for the search of the BEDs begins. During the search, we continue to monitor the toxicity rates and update the MTD. In this stage, only the admissible doses, those not exceeding the current MTD, are considered in dose assignments. Because dose-finding trials typically have small sample sizes, the efficacy rate for each individual dose cannot be reliably estimated; therefore, we group the adjacent doses to obtain pooled estimates. Afterward, we approximate the unimodal dose–efficacy curve by a step function that divides the dose axis into three intervals/groups as illustrated in Fig. 1. The middle interval of the step function which maximizes a utility function will be recommended as the interval of BEDs. After the interval of BEDs has been recommended, a confirmation cohort is added to the trial to increase the precision of the

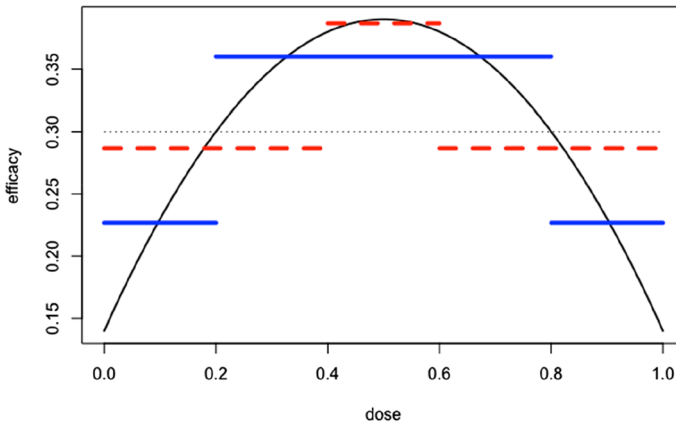


Fig. 1 Plot of an efficacy curve with the minimum acceptable rate 0.30 and two-step function models: the best one in blue and another one in red

estimate of the pooled efficacy rate of this interval. The point estimate, together with confidence intervals of the efficacy rate for each BED and of the pooled efficacy rate for all BEDs, can be provided to help clinicians select a dose to move forward to phase II trials.

2.1 Modelling Toxicity and Efficacy Rates

Let the dose levels be indicated in discrete integers i for $i = 1, \dots, m$. Let $P_{T,i}$ and $P_{E,i}$ be the toxicity and efficacy rates, respectively, of dose i . We assume that the toxicity rates are monotonically non-decreasing, that is, $P_{T,1} \leq P_{T,2} \leq \dots \leq P_{T,m}$; it follows that all doses not exceeding the current estimated MTD are admissible. The prior distribution of the toxicity rate at each dose level is taken to be independent beta distributions. Specifically, we have $P_{T,i}$ independently distributed as $\text{Beta}(\alpha_{T,i}^0, \beta_{T,i}^0)$ such that the means are monotonically non-decreasing: $\mu_{T,1}^0 \leq \mu_{T,2}^0 \leq \dots \leq \mu_{T,m}^0$, where $\mu_{T,i}^0 = \alpha_{T,i}^0 / (\alpha_{T,i}^0 + \beta_{T,i}^0)$. Note that the monotonically non-decreasing property will be preserved in the posterior means. Throughout, the superscript 0 is used to indicate that a parameter is associated with an initial prior distribution.

While it is reasonable to assume that the toxicity of an agent is monotonic and non-decreasing in dose, this assumption may not be tenable for the efficacy [6]. Therefore, we assume instead that the efficacy curve $f(s)$ is unimodal or plateaued: $f(s)$ increases at low dose levels and plateaus or even decreases at higher levels. It follows that the set of BEDs is an interval. Because our goal is to identify the interval of BEDs, rather than the BOD, we model the unimodal or plateaued efficacy curve using a step function, which is identified by an interval $B = [l, r]$, $1 \leq l \leq r \leq m$ as follows:

$$P_{E,i} = \begin{cases} \frac{1}{l-1} \sum_{j=1}^{l-1} P_{E,j} & \text{for } i < l, \\ \frac{1}{r-l} \sum_{j=l}^r P_{E,j} & \text{for } l \leq i \leq r, \\ \frac{1}{m-r-1} \sum_{j=r+1}^m P_{E,j} & \text{for } i > r. \end{cases} \tag{1}$$

The efficacy rates of doses in the same range: to the left of, to the right of, or in the interval B , are modelled as the mean efficacy rate of all doses in the range.

The prior distribution of the efficacy rate at each dose level is taken to be independent beta distributions, that is, we have $P_{E,i}^0$ independently distributed as $\text{Beta}(\alpha_{E,i}^0, \beta_{E,i}^0)$ for $i=1, \dots, m$. Similarly, we specify the initial prior distribution of the mean efficacy rate of an interval A of k doses, $P_A = \frac{1}{k} \sum_{i \in A} P_{E,i}$, as $\text{Beta}(\alpha_A^0 = \frac{1}{k} \sum_{i \in A} \alpha_{E,i}^0, \beta_A^0 = \frac{1}{k} \sum_{i \in A} \beta_{E,i}^0)$. We defer the discussion of the specification of the prior parameters to Sect. 3.

2.2 Utility Functions

Let T_{\max} be the maximum tolerable toxicity level and E_{\min} be the minimum acceptable efficacy rate such that a BED should have its efficacy rate above it. Doses with tolerable toxicity and acceptable efficacy are BEDs. The goal of our design is to identify the interval of BEDs correctly and promptly. For an arbitrary interval $B = [l, r]$, its utility of being the interval of BEDs is defined as follows.

Considering a dose i , if it is a BED, it should be included in B ; if it is a non-BED, it should be excluded from B . Among all BEDs, the successful inclusion of a more efficacious dose should be given larger utility than that of a less efficacious dose. In contrast, among all non-BEDs, the successful exclusion of a less efficacious dose should be given larger utility than that of a more efficacious dose. The utility of interval $B = [l, r]$ associated with dose i is defined as follows:

$$u_i(B) = \begin{cases} 1_{\{P_{E,i} \geq E_{\min}\}} [a + b(P_{E,i} - E_{\min})] & \text{for } i \in B \\ 1_{\{P_{E,i} < E_{\min}\}} [c + d(E_{\min} - P_{E,i})] & \text{for } i \notin B \end{cases} \tag{2}$$

where $1_{\text{condition}}$ takes the value one if *condition* is true and zero otherwise. The positive constants a and b , which are associated with the BEDs, represent the weights for correct selection as BEDs and how much more efficacious it is than E_{\min} , respectively. The positive constants c and d , which are associated with the non-BEDs, represent the weights for correct selection as non-BEDs and how much less efficacious it is than E_{\min} , respectively. The total utility is the weighted sum of the utilities at all doses:

$$u(B) = \sum_{i=1}^{\text{MTD}} w_i * u_i(B). \tag{3}$$

An intuitive choice for the weight w_i is the sample size at dose i , denoted n_i , but it can be chosen in other ways.

The utility function $u(B)$ is maximized when the interval B contains all and only BEDs. For any BED, say dose i , $u_i(B)$ is at least one if it is in B and zero if not.

Similarly, for any non-BEDs, say dose j , $u_j(B)$ is zero if it is in B and at least one if not. To maximize the total sum of the utilities at all doses, the interval B must include all BEDs and exclude all non-BEDs.

When a prior distribution is added on $P_{E,i}$, the (Bayesian) utility associated with dose i is defined as the posterior expectation of $u_i(B)$, $E[u_i(B)]$. Let $n_{e,i}$ be the number of efficacy outcomes among the n_i patients treated at dose i . The posterior distribution of $P_{E,i}$ is Beta($\alpha_{E,i} = \alpha_{E,i}^0 + n_{e,i}$, $\beta_{E,i} = \beta_{E,i}^0 + n_i - n_{e,i}$) and the expected utility at dose i is given by:

$$E[u_i(B)] = \begin{cases} (a - b^*E_{\min}) [1 - I_{E_{\min}}(\alpha_{E,i}, \beta_{E,i})] + \frac{b\alpha_{E,i}}{\alpha_{E,i} + \beta_{E,i}} [1 - I_{E_{\min}}(\alpha_{E,i} + 1, \beta_{E,i})] & \text{for } i \in B, \\ (c + d^*E_{\min}) I_{E_{\min}}(\alpha_{E,i}, \beta_{E,i}) - \frac{d\alpha_{E,i}}{\alpha_{E,i} + \beta_{E,i}} I_{E_{\min}}(\alpha_{E,i} + 1, \beta_{E,i}) & \text{for } i \notin B, \end{cases}$$

where $I_{E_{\min}}(\bullet, \bullet)$ is the regularized incomplete beta function. It follows that there is an analytical expression for the expected total utility of an interval B :

$$E[u(B)] = \sum_{i=1}^{MTD} w_i^* E[u_i(B)]$$

Let A^* denote the interval which maximizes $E[u(B)]$ at the end of trial. Due to limited sample size, not all doses in A^* are necessarily BEDs and so we need to exclude the non-BEDs as follows. First of all, because of the assumed unimodality of the dose–efficacy curve, any non-BEDs in the interval A^* must be the lower or higher dose levels in A^* . Let B be a subinterval of A^* obtained by excluding one or more lower and/or higher dose levels in A^* , and let k be the number of doses in B . Next, we define the efficacy rate of B as the pooled efficacy rate $P_B = \frac{1}{k} \sum_{i \in B} P_{E,i}$, and we specify the initial prior distribution of P_B as the pooled beta prior distribution Beta($\alpha_B^0 = \frac{1}{k} \sum_{i \in B} \alpha_{E,i}^0$, $\beta_B^0 = \frac{1}{k} \sum_{i \in B} \beta_{E,i}^0$). Upon the collection of data, the posterior distribution of P_B becomes Beta($\alpha_B = \alpha_B^0 + \sum_{i \in B} n_{e,i}$, $\beta_B = \beta_B^0 + \sum_{i \in B} (n_i - n_{e,i})$), and the posterior probability that B is acceptable is given by $P(P_B \geq E_{\min}) = 1 - I_{E_{\min}}(\alpha_B, \beta_B)$. Our design recommends the subinterval $B^* = [l^*, r^*]$ that maximizes this posterior probability as the final interval of BEDs.

2.3 Dose Selection Rules

To use the sample size more efficiently, stopping rules are implemented at both stages. Let $n_{\min.mtd}, n_{\max.mtd}$ be the minimum and maximum sample sizes at the first stage for locating the MTD, and n_{\min}, n_{\max} be the minimum and maximum samples sizes of the entire trial. In addition, $n_{\min.mtd} \leq n_{\max.mtd} \leq n_{\min} \leq n_{\max}$.

At the beginning of a trial, phase Ia, we propose the following dose selection rules to locate the MTD quickly:

- (1) Start at the lowest dose and escalate until a toxicity outcome is observed or the highest dose is reached.

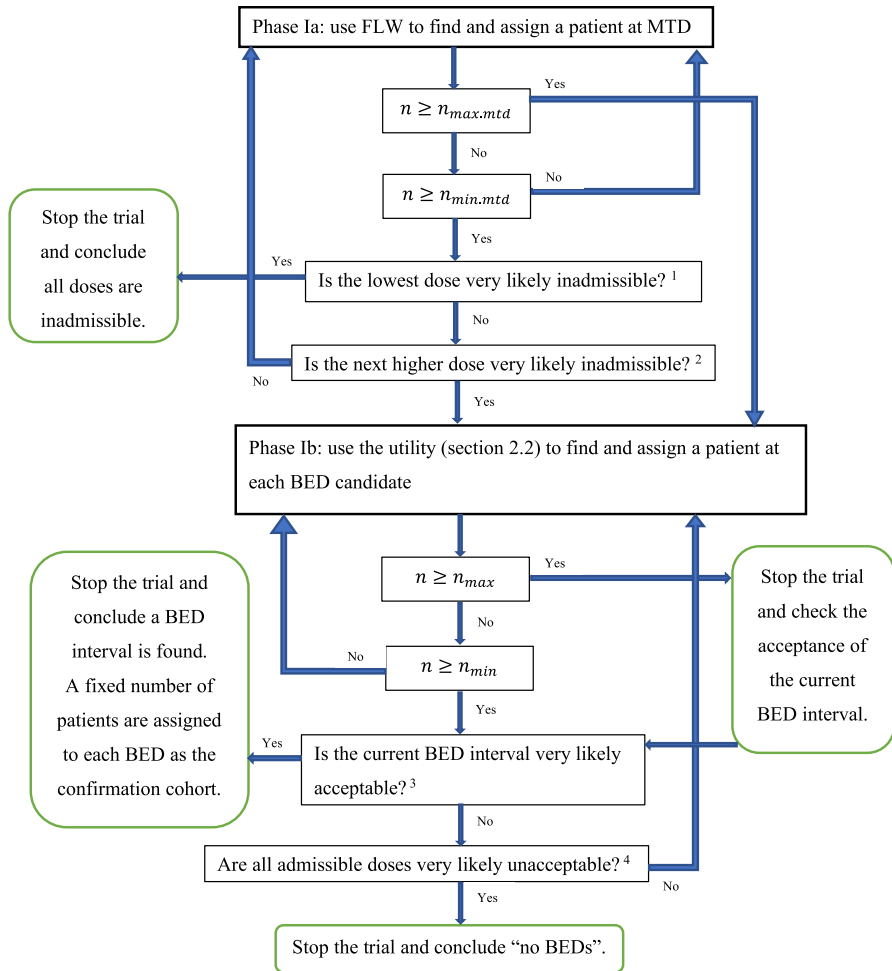
- (2) Use the FLW algorithm to locate the MTD and assign the next patient to the current MTD.
- (3) Step (2) is repeated until a pre-set sample size $n_{\min.mtd}$ is reached. The dose assignment between sample size $n_{\min.mtd}$ and $n_{\max.mtd}$ follows the rules below:
 - (a) If the posterior probability of the first dose being overly toxic exceeds a threshold, the trial is terminated and we conclude that the MTD is below the first dose.
 - (b) Otherwise, if the posterior probability of the current MTD being the true MTD exceeds a threshold, we will begin to allocate the BED candidates to the patients.
 - (c) Otherwise, the current MTD is assigned to the next patient.

The parameter $n_{\min.mtd}$ in Step (3) is designed to prevent incorrect early termination due to misspecified prior distributions (the conclusion “all doses are inadmissible” cannot be made before $n_{\min.mtd}$ patients are treated), while the parameter $n_{\max.mtd}$ is designed to ensure possible allocation of the BEDs to patients. One can set $n_{\min.mtd} = n_{\max.mtd}$ to simplify the algorithm. The reason to allocate the MTD rather than BEDs at the beginning of the trial is to identify the MTD quickly, since patient safety should always be the first priority. Once the MTD is confidently located, phase Ib begins and patients will be allocated to BED candidates immediately.

The BED candidates are defined to be all the admissible doses in the interval of BEDs, $B^* = [l^*, r^*]$ (as described in the end of Sect. 2.2) and its adjacent doses $l^* - 1$ and $r^* + 1$. To ensure patient safety, only admissible doses at or below the current MTD can be assigned to patients, and the MTD is updated throughout the entire trial using the FLW algorithm and observed data at all doses. Let q_1 be the pres-set threshold of the posterior probability of a dose being acceptable and q_2 be the threshold of the posterior probability of a dose being unacceptable. Each BED candidate is assigned to a patient until a minimum sample size n_{\min} is reached and the following stopping rules apply:

- (1) If the current interval of BEDs, B^* , is very likely to be acceptable, that is, $P(P_{B^*} \geq E_{\min}) \geq q_1$, the trial is terminated and we recommend B^* as the interval of BEDs.
- (2) Otherwise, if all admissible doses are very likely to be unacceptable, that is, $P(P_{E,i} < E_{\min}) \geq q_2$ for all $i \leq \text{MTD}$, the trial is terminated and we conclude that there is no BED.
- (3) Otherwise, the trial is terminated until the maximum sample size n_{\max} is reached. If the current interval of BEDs, B^* , is very likely to be unacceptable, that is, $P(P_{B^*} < E_{\min}) \geq q_2$, we conclude that there is no BED; otherwise, we recommend B^* as the interval of BEDs.

A flowchart of the proposed phase Ia/Ib design is presented in Fig. 2.



1. $P(P_{T,1} \geq T_{max}) \geq q_1$, where $P_{T,1}$ is the toxicity rate of the lowest dose, T_{max} is the maximum tolerable toxicity rate, and q_1 is a pre-set threshold such as .90.
2. $P(P_{T,MTD+1} \geq T_{max}) \geq q_1$, where $P_{T,MTD+1}$ is the toxicity rate of the next higher dose of the current MTD.
3. $P(P_A \geq E_{min}) \geq q_1$, where P_A is the pooled toxicity rate of the current BED interval A , E_{min} is the minimal acceptable efficacy rate.
4. $P(P_{E,i} < E_{min}) \geq q_2$ for all $i \leq \text{MTD}$, where $P_{E,i}$ is the efficacy rate of dose i and q_2 is a pre-set threshold such as .90.

Fig. 2 Flowchart of CFBD

2.4 Confirmation Cohort

If at the end of trial, an interval $B^* = [l^*, r^*]$ is recommended as the BED interval, a confirmation cohort will be recruited to improve the efficiency of estimates. Each dose in the interval B will be assigned the same number of patients and this

cohort is later referred to the confirmation cohort. After the confirmation cohort, the selection of the MTD and BED interval will be updated. In addition, the pooled efficacy rate for the updated interval $B^* = [l^*, r^*]$ is empirically estimated by $\widehat{P}_{B^*} = \sum_{i=l^*}^{r^*} n_{e,i} / \sum_{i=l^*}^{r^*} n_i$. The pooled efficacy rate is adopted rather than individual efficacy rate because of the small sample size at each individual dose. The size of the confirmation cohort depends on the desired efficacy and available resources. However, to detect a small difference from the target efficacy level E_{\min} , it requires a large sample size at these BEDs. In addition, the goal of dose finding trials is to locate the good acceptable doses, not the precision of estimation. The confirmation cohort is only used to select one of these BEDs as the single dose to move to phase II trials.

3 Simulation Studies

Hereafter, we refer to our proposed design as CFBD, which stands for Curve-Free Bayesian Design. The CFBD is not the only curve-free Bayesian design in the literature, for example, Zack and Lee (2017) also proposed a curve-free Bayesian design (ZL) to identify the BOD. In our simulation studies, we compare the performance of CFBD to ZL because there are limited number of designs in the literature that seek to identify the BEDs, and ZL uses the same two-stage strategy and a similar curve-free Bayesian model. We would like to emphasize that the goals of CFBD and ZL are different and, as a result, the CFBD may outperform ZL in some scenarios simply because ZL was designed to identify the BOD rather than BEDs. In other words, our comparisons are not intended as a demonstration of the superiority of the CFBD; rather, the performance of ZL when used to identify the set of BEDs serves only as a benchmark.

3.1 Operating Characteristics

3.1.1 Simulation Settings

We examined the operating characteristics of the CFBD in six scenarios presented in Zang and Lee [18]. These scenarios, reproduced in Table 1 with the BEDs highlighted in bold, provide a variety of dose-efficacy curves: monotone increasing (scenarios 1 and 6), unimodal (scenarios 2 and 3), and increasing with a plateau (scenarios 4 and 5). Note that a step function would fit the dose-efficacy curve well in scenarios 4 and 5 but not apparently so in other scenarios. We simulated 5000 trials for each scenario, where the value of both the maximum tolerable toxicity rate, T_{\max} , and the minimum acceptable efficacy rate, E_{\min} , was 0.3.

In CFBD, we proposed the Beta distributions as the prior distributions of the toxicity and efficacy rates. Our choice of the parameters in the simulation studies was motivated by the following considerations. First of all, the values of the parameters should not be large, since a $\text{Beta}(\alpha, \beta)$ prior distribution has effective sample size $r = \alpha + \beta$ [12]. If $r > n$, where n is the sample size, the prior dominates posterior inferences, and vice versa. Therefore, unless there is reliable prior information, r

Table 1 Six scenarios with its shape of efficacy curve; doses in bold are BEDs

Scenario		Dose level				
		1	2	3	4	5
1	Toxicity	0.10	0.20	0.30	0.40	0.50
Increasing	Efficacy	0.05	0.10	0.18	0.25	0.30
2	Toxicity	0.01	0.05	0.09	0.15	0.20
Unimodal	Efficacy	0.10	0.30	0.40	0.20	0.05
3	Toxicity	0.02	0.06	0.12	0.30	0.50
Unimodal	Efficacy	0.30	0.40	0.20	0.10	0.05
4	Toxicity	0.02	0.06	0.12	0.30	0.50
Plateau	Efficacy	0.10	0.30	0.30	0.30	0.30
5	Toxicity	0.10	0.20	0.25	0.40	0.50
Plateau	Efficacy	0.20	0.40	0.40	0.40	0.40
6	Toxicity	0.05	0.10	0.20	0.35	0.50
Increasing	Efficacy	0.05	0.20	0.30	0.40	0.50

should be fairly small so that any misspecified priors can be correctly updated with a small or moderate sample that is typical in Phase I trials. Secondly, a Beta(α, β) distribution is unimodal if and only if $\alpha > 1$ and $\beta > 1$. As unimodal prior distributions are usually preferred, the parameters α^0 and β^0 of the Beta prior distributions should be larger than 1 if possible. Combining these two considerations, we should have α^0 and β^0 larger than but close to 1. However, this condition may not be satisfied at all dose levels, and so we require instead that it be at least satisfied at dose levels that are believed to be BEDs prior to the trial. Let $\mu_{E,i}^0$ and r_i^0 denote the mean and effective sample size, respectively, of a Beta($\alpha_{E,i}^0, \beta_{E,i}^0$) prior distribution for the efficacy rate at dose i . It is trivial to show that $\alpha_{E,i}^0 = r_i^0 \mu_{E,i}^0$ and $\beta_{E,i}^0 = r_i^0 (1 - \mu_{E,i}^0)$, so that the equivalent condition is to have both $r_i^0 \mu_{E,i}^0$ and $r_i^0 (1 - \mu_{E,i}^0)$ greater than but close to 1 at the likely BEDs. In practice, these parameters can be specified in collaboration with clinicians based on their clinical experience and confidence level. Suppose that all efficacy rates of the BEDs are expected to be above E_{\min} but below an upper bound of $E_{\max} (\geq E_{\min})$. Let r be a multiplier such that $rE_{\min}, r(1 - E_{\min}), rE_{\max}$ and $r(1 - E_{\max})$ are all larger than but close to 1. Then $\alpha_{E,i}^0 = r\mu_{E,i}^0$ and $\beta_{E,i}^0 = r(1 - \mu_{E,i}^0)$ ensure that the prior distributions are unimodal whenever $E_{\min} \leq \mu_{E,i}^0 \leq E_{\max}$. For now, $\mu_{E,i}^0$ is chosen to be equal to the actual efficacy rate at dose i ; we will study the impact of misspecifications in Sect. 3.2. With $E_{\min} = 0.3$ and $E_{\max} = 0.4$ in our simulations, we have the effective sample size $r = 4$. Based on our earlier discussion, our prior information is as strong as that obtained from four treated patients.

The weights a, b, c , and d , which represent the rewards in the utility function defined in Eq. (2), are all set to 1, and the weights w_i used in the calculation of the total utility defined in Eq. (3) are given by the sample sizes, n_i , at the respective dose levels. The parameters of stopping rules for the simulated trials are given as follows. A trial should not be stopped when the sample size is small. The

minimal and maximal sample size for MTD identification (phase Ia) are set as $n_{\min.mtd} = 10$, and $n_{\max.mtd} = 30$. In addition, it takes about 12 patients to have a power of 0.70 to test the hypothesis $H_0: P_{E,i} = E_{\min} = 0.3$ against $H_1: P_{E,i} = 0.1$ at a significance level of 0.1. Therefore, with 5 doses in our simulated trials, we set the minimum sample size to stop a trial early as $n_{\min.int} = 5 \times 12 = 60$. Although Zang and Lee [18] set the maximum sample size, $n_{\max.int}$, as 150, we chose 100 instead to reflect the sample size constraint of actual phase I trials and reserved the remaining 50 for possible confirmation cohorts later. The size of the confirmation cohort per BED is chosen to be 10. Finally, the thresholds q_1 and q_2 for stopping rules defined in Sect. 2.3 are set to be 0.9.

3.1.2 Simulation Results

The simulation results are summarized in Tables 2 and 3. Recall that Zang and Lee [18] aim to identify the BOD (the BED with the highest efficacy rate) whereas CFBD aims to identify all BEDs. To facilitate comparisons, we compute the following summary statistics in Table 2:

- The percentage of trials that conclude that there is a target dose (%found), which is defined for CFBD as the percentage of trials recommending BEDs, and for ZL as the percentages of trials recommending a BOD.

Table 2 The average sample size (\bar{n}), the percentage of trials that recommend any BEDs/ BOD (%found), within the trials recommending BEDs/ BOD, the percentage that all the recommended doses are truly acceptable (%correct) and the percentage that sample pooled efficacy of the BEDs are acceptable (%success), and the percentages of in-trial toxicity (%tox) and efficacy (%eff) under the proposed Bayesian design (CFBD) and the design proposed by Zang and Lee (ZL)

Design	\bar{n}	%found	%correct	%success	%tox	%eff
Scenario 1: increasing efficacy curve, no BEDs						
CFBD	67.0	6.6	0.0	17.7	20.6	11.3
ZL	30.4	2.6	0.0		26.3	9.9
Scenario 2: unimodal efficacy curve, two BEDs in middle						
CFBD	90.7	99.2	96.7	93.7	10.2	26.2
ZL	114.8	94.6	88.7		7.9	28.4
Scenario 3: unimodal efficacy curve, two BEDs in front						
CFBD	89.2	99.7	97.2	95.3	9.0	29.2
ZL	110.0	90.9	95.9		8.2	31.5
Scenario 4: step efficacy curve, three BEDs in middle						
CFBD	99.4	96.9	98.3	81.3	14.1	27.5
ZL	108.6	91.2	87.7		11.9	28.0
Scenario 5: step efficacy curve, two BEDs in middle						
CFBD	83.0	94.2	90.3	97.6	20.2	34.7
ZL	80.6	75.5	75.2		23.8	29.8
Scenario 6: increasing efficacy curve, one BED in middle						
CFBD	92.1	80.5	48.5	75.1	18.3	26.3
ZL	90.5	64.4	45.5		19.0	23.9

Table 3 The selection percentages of MTD, BED interval limits: L^* , U^* , and patient allocation for all doses under CFBD; correct percentages are in bold

Dose level	1	2	3	4	5
Scenario 1: increasing efficacy curve, no BEDs					
% of MTD	6.1	48.4	40.7	4.8	0.1
% of L^*	2.8	0.5	2.0	1.3	0.0
% of U^*	2.8	0.5	2.0	1.3	0.0
% of patients	31.4	40.1	21.1	7.0	0.5
Scenario 2: unimodal efficacy curve, two BEDs in middle					
% of MTD	0.0	0.7	8.6	32	58.6
% of L^*	0.5	27.0	70.5	1.3	0.0
% of U^*	0.5	10.7	85.2	2.8	0.0
% of patients	7.8	24.7	32.6	21.0	13.9
Scenario 3: unimodal efficacy curve, two BEDs in front					
% of MTD	0.1	2.8	26.6	68.7	1.8
% of L^*	25.2	73.3	1.1	0.1	0.0
% of U^*	8.4	88.4	2.7	0.1	0.0
% of patients	28.7	36.6	24.9	8.3	1.4
Scenario 4: step efficacy curve, three BEDs in middle					
% of MTD	0.3	2.4	25.2	71.3	0.9
% of L^*	1.3	42.1	32.8	20.6	0.1
% of U^*	1.3	32.1	36.0	27.1	0.3
% of patients	12.4	26.6	35.2	24.2	1.6
Scenario 5: step efficacy curve, two BEDs in middle					
% of MTD	6.0	37.8	45.3	10.9	0.0
% of L^*	2.2	74.3	16.0	1.7	0.0
% of U^*	1.1	50.6	35.5	7.0	0.0
% of patients	26.7	40.0	25.7	7.2	0.4
Scenario 6: increasing efficacy curve, one BED in middle					
% of MTD	0.4	9.5	56.9	33.0	0.2
% of L^*	3.4	8.7	44.8	23.5	0.1
% of U^*	3.4	7.9	39.5	29.4	0.2
% of patients	10.2	30.6	41.0	17.5	0.7

- Within those trials recommending BEDs/BOD, the percentage of trials of which all recommended doses are truly admissible and acceptable (%correct).
- Within the trials recommending BEDs, the percentage of trials of which the sample pooled efficacy rate of the recommended BEDs is acceptable, i.e. confidence level of the pooled efficacy rate being above E_{\min} (%success). This statistic is computed for CFBD only.
- The percentage of in-trial toxicities (%tox), the percentage of in-trial efficacies (%eff), and the average sample size over the 5000 simulated trials (\bar{n}).

Additional summary statistics are provided in Table 3.

As seen from Table 2, CFBD performs very well in identifying the BEDs in most scenarios that have at least one BED. In scenarios 2 to 5, CFBD uses relatively fewer patients on average than ZL and yet has higher %found and %correct. Also, CFBD was able to recommend one or more BEDs in approximately 95% or more of the trials (%found) with at least 90% accuracy rate (%correct). In scenario 6, CFBD was able to detect the existence of BEDs 80.5% of the time, which is 16% more likely than ZL; however, the recommended doses are not very likely (less than 50%) to be all acceptable. An explanation for the low %correct is the existence of a boundary dose. Dose 4 has a high efficacy rate (0.4 compared to $E_{\min}=0.3$) and a toxicity rate that is close to being admissible (0.35 compared to $T_{\max}=0.3$). Due to the small sample size at dose 4, the toxicity rate of 0.35 cannot be distinguished from 0.30 and so CFBD often recommended dose 4 as a BED in the simulated trials. More specifically, dose 4 was often selected (29.4% of times) as the upper limit, U^* , of the recommended interval of BEDs (see Table 3). If a toxicity rate of 0.35 is considered admissible, the percentage of trials that provided a correct recommendation would increase dramatically. It is observed that for all scenarios with BEDs (scenarios 2 to 6), the confidence level of the recommended BEDs being acceptable as a group (%success) is different from the true percentage of the recommended BEDs being all acceptable (%correct). This issue can be fixed by increasing the sample size in the confirmation cohort. However, the precision of efficacy estimation is not of interest in phase I trials and this issue can be deferred to phase II trials.

In scenario 1, where there is no BED, the percentage of trials in which CFBD did not recommend any dose ($100 - \%found$) is comparable to that of ZL (93.4% versus 97.4%), but the former tended to stop the trials much later than the latter (an average of 67.0 versus 30.4 patients). This is due to the stopping rule $n_{\min.int}=60$ that is imposed on CFBD. If $n_{\min.int}=10$, the average sample size for CFBD drops to 26.7 while the percentage of trials in which CFBD did not recommend an interval remained at the high level (94.6%).

Overall, CFBD performs wells in terms of the average sample size and the percentage of good recommendation, regardless of the shape of the underlying efficacy curve.

3.2 Sensitivity Analysis

We conducted a sensitivity analysis to investigate the robustness of CFBD. There were two sets of simulations, one involving random errors and the other involving fixed errors in prior specifications. In the first set of simulations, we added random errors to the means of the Beta prior distributions, which were originally chosen to be equal to the true rates. Specifically, we set initial Beta mean toxicity and efficacy rate $\mu_{T,i}^0 = (1 + \epsilon_{T,i})P_{T,i}$ and $\mu_{E,i}^0 = (1 + \epsilon_{E,i})P_{E,i}$, where $\epsilon_{T,i}$ and $\epsilon_{E,i}$ are uniformly distributed between -50% and 50% of the corresponding true rates. All other settings were as described in the previous section. Table 5 summarizes the results, from which we see that CFBD is extremely robust: the operating characteristic of CFBD under misspecifications of parameter values, with random errors up to 50% of the true values, is nearly identical to that without misspecification. We

attribute this robustness to adequate sample size, relatively weak prior information, and the curve-free approach. Recall that the minimum sample size of 60 patients was selected based on the consideration of power and significance level of a test of $H_0: P_{E,I} = E_{\min}$ (Sect. 3.1). Because every admissible dose in and around the interval of BEDs, rather than only the BOD, were assigned to the next cohort of patients, a given sample of patients would be fairly uniformly distributed among the admissible doses, such that an average of 12 or more patients were assigned to each admissible dose. Also recall that the parameters of the prior distributions were chosen such that the prior information is as strong as that obtained from four patients, translates into one-third of the average number of patients assigned to each admissible dose. This allows the data to correctly update the misspecified priors. Finally, without a parametric assumption on the dose–toxicity and dose–efficacy relationships, the data can freely correct any errors in the priors.

To investigate the robustness of CFBD to extremer misspecifications, we modified the original toxicity and efficacy scenarios in the second set of simulations. The (prior) toxicity rates were shifted such that the prior MTD is different from the true MTD, and the (prior) efficacy rates were also shifted or modified such that the prior set of BEDs and the true set of BEDs are disjoint (Table 4). All other settings remained the same as described in Sect. 3.1. The results are summarized in Table 5. Although the priors were very different from the true scenario, and the prior MTD and BEDs were completely different from the actual values, CFBD still performed reasonably well. In scenario 1, the only scenario with no BED, CFBD made the correct conclusion in 86.7%

Table 4 Sensitivity analysis: fixed error priors with incorrect BED sets (disjoint with the true BED sets)

Dose level	True rate (prior rate)				
	1	2	3	4	5
Scenario 1: no BED becomes one BED					
Toxicity	0.10 (0.05)	0.20 (0.10)	0.30 (0.20)	0.40 (0.30)	0.50 (0.40)
Efficacy	0.05 (0.10)	0.10 (0.18)	0.18 (0.25)	0.25 (0.30)	0.30 (0.40)
Scenario 2: BED set {2,3} becomes {4}					
Toxicity	0.01 (0.09)	0.05 (0.15)	0.09 (0.20)	0.15 (0.30)	0.20 (0.40)
Efficacy	0.10 (0.05)	0.30 (0.10)	0.40 (0.20)	0.20 (0.30)	0.05 (0.40)
Scenario 3: BED set {1,2} becomes {3}					
Toxicity	0.02 (0.06)	0.06 (0.12)	0.12 (0.30)	0.30 (0.50)	0.50 (0.60)
Efficacy	0.30 (0.05)	0.40 (0.10)	0.20 (0.30)	0.10 (0.40)	0.05 (0.20)
Scenario 4: BED set {2,3,4} becomes {1}					
Toxicity	0.02 (0.12)	0.06 (0.35)	0.12 (0.50)	0.30 (0.55)	0.50 (0.60)
Efficacy	0.10 (0.30)	0.30 (0.30)	0.30 (0.30)	0.30 (0.30)	0.30 (0.10)
Scenario 5: BED set {2,3} becomes {4}					
Toxicity	0.10 (0.05)	0.20 (0.10)	0.25 (0.20)	0.40 (0.25)	0.50 (0.40)
Efficacy	0.20 (0.20)	0.40 (0.20)	0.40 (0.20)	0.40 (0.40)	0.40 (0.40)
Scenario 6: BED set {3} becomes {2}					
Toxicity	0.05 (0.10)	0.10 (0.20)	0.20 (0.35)	0.35 (0.50)	0.50 (0.55)
Efficacy	0.05 (0.20)	0.20 (0.30)	0.30 (0.40)	0.40 (0.50)	0.50 (0.55)

Table 5 Sensitivity analysis: average sample size (\bar{n}), the percentage of trials that recommended any BEDs (%found), within the trial found BEDs, the percentage that recommended BEDs are acceptable (%correct) and the percentage that data suggest acceptable pooled efficacy (%success), and the percentages of in-trial toxicity (%tox) and efficacy (%eff) under correct priors, incorrect priors with random errors and with fixed errors

Prior type	\bar{n}	%found	%correct	%success	%tox	%eff
Scenario 1: increasing efficacy curve, no BEDs						
Correct	67.0	6.6	0.0	17.7	20.6	11.3
Random error	67.5	6.8	0.0	21.5	20.9	11.4
Fixed error	72.4	13.3	0.0	24.5	23.5	13.1
Scenario 2: unimodal efficacy curve, two BEDs in middle						
Correct	90.7	99.2	96.7	93.7	10.2	26.2
Random error	90.4	98.9	96.5	94.1	10.2	26.0
Fixed error	94.9	97.5	95.7	93.9	10.1	26.9
Scenario 3: unimodal efficacy curve, two BEDs in front						
Correct	89.2	99.7	97.2	95.3	9.0	29.2
Random error	89.2	99.5	97.4	95.4	9.3	29.1
Fixed error	96.8	98.4	96.6	95.3	8.5	29.4
Scenario 4: step efficacy curve, three BEDs in middle						
Correct	99.4	96.9	98.3	81.3	14.1	27.5
Random error	99.1	96.6	98.4	81.2	14.2	27.5
Fixed error	98.8	90.3	97.5	73.3	8.3	25.6
Scenario 5: step efficacy curve, two BEDs in middle						
Correct	83.0	94.2	90.3	97.6	20.2	34.7
Random error	83.0	94.6	91.4	96.9	20.1	34.6
Fixed error	87.0	97.1	83.4	97.7	22.7	36.0
Scenario 6: increasing efficacy curve, one BED in middle						
Correct	92.1	80.5	48.5	75.1	18.3	26.3
Random error	91.9	80.6	48.3	75.4	18.6	26.5
Fixed error	91.1	70.3	60.2	62.6	15.0	22.9

of simulated trials (100 – %found). For the scenarios with BEDs, CFBD was able to recommend one or more BEDs in 90% or more of the trials (%found) in scenarios 2 to 5, with at least 83% accuracy rate (%correct). In scenario 6, where there is a boundary dose (dose 4), CFBD under misspecified priors actually made a good recommendation (%found * %correct) slightly more often than CFBD under correct priors (42% vs. 39%). The reason is, the misspecified priors overestimated both toxicity and efficacy rates, which rendered the boundary dose inadmissible in more trials and thereby increased %correct, the percentage of trials that provided a correct recommendation within those trials recommending BEDs. Overall, in our simulation studies, CFBD still remains pretty robust with slightly larger sample size (less than 10 more on average) even with such greatly misspecified prior.

4 A Demonstration Based On Simulations Using Real Trial Data

A completed phase I trial conducted at the University of California at Los Angeles Medical Center [13] was used to demonstrate an application of the proposed CFBD. The purpose of the trial is to recommend the BOD of several doses of celecoxib combined with erlotinib at a fixed dose in patients diagnosed with advanced non-small-cell lung cancer (NSCLC). Celecoxib is a cyclooxygenase-2 (COX-2) inhibitor, which can regulate cellular proliferation, migration, and invasion. On the other hand, erlotinib is a highly specific epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI), which has been shown to inhibit the growth of human cancer cells. Several studies by 2006, for example Kryan et al. [9], found that overexpression of COX-2 promotes EGFR-TKI resistance. The knowledge of COX-2 and EGFR in NSCLC and the interaction of their signalling pathways provided a unique way to inhibit tumour angiogenesis, invasion, and growth. Motivated by these findings, Reckamp et al. [13] conducted a combined trial to establish a BOD within 200, 300, 400, 600 and 800 mg of celecoxib administered with 150 mg of erlotinib. The BOD was determined at the lowest dose level (of celecoxib) showing optimal biological activity, defined as a maximal decrease in the level of urinary prostaglandin E-M (PGE-M), where no dose-limiting toxicity (DLT) occurred.

Twenty-two subjects were enrolled and 21 were evaluable for the determination of the BOD, toxicity and efficacy. There were no DLTs among the 21 subjects and thus a low toxicity profile is assumed. The adopted toxicity profile was modified slightly based on scenario 2, the lowest toxicity scenario in Table 1 in our previous simulation studies. The efficacy outcome was defined as a decline in urinary PGE-M of $> 72\%$, which is known to improve overall survival. Based on the data of three patients per dose level, the sample mean of the decline percentages in urinary PGE-M at the five dose levels were around 9%, 8%, 65%, 84%, and 87%, respectively, with a sample standard deviation of less than 10%. In our illustration of CFBD, we assume the decline percentage follows a normal distribution with the mean 10%, 15%, 70%, 80%, 80%, and a standard deviation of 15% for all dose levels. The enlarged standard deviation and slightly different means are intended to reflect the uncertainty due to the small sample. The true efficacy rate, a probability of more than 72% decline in the urinary PGE-M, can then be calculated. Furthermore, the prior efficacy rates are based on the Emax curve [11]. The efficacy profile together with the toxicity profile is reported in Table 6.

Different to Reckamp's phase I/II trial, our method is designed for phase Ia/Ib trials and our goal is to find all BEDs, that is, all admissible doses with sufficient efficacy. Because the sample size of Reckamp's trial was much smaller than that used in the simulation studies presented in Sect. 3.1.2, we do not recruit any confirmation cohorts in the current simulated trials. In addition, to reflect the lower toxicity and higher efficacy rates in the actual trial, we set the maximal tolerable DLT rate as $T_{\max} = 0.20$ and the minimal acceptable efficacy rate as $E_{\min} = 0.40$. All other parameters remain the same as in Sect. 3. The summary statistics over 5000 simulated trials are reported in Table 6. The results show that CFBD is able

Table 6 Demo trial: the selection percentages of MTD, BED interval limits: L^* , U^* , and patient allocation for all doses; average sample size (\bar{n}), percentages of none and good BED selection (%none, %good), in-trial toxicity and efficacy rates (%tox, %eff); MTD and BED values are in bold

Dose level	200 mg	300 mg	400 mg	600 mg	800 mg
True toxicity	0.00	0.01	0.05	0.09	0.15
Prior toxicity	0.00	0.01	0.05	0.09	0.15
True efficacy	0.00	0.00	0.45	0.70	0.70
Prior efficacy	0.00	0.09	0.40	0.78	0.79
% of MTD	0.0	0.1	4.3	25.6	70.0
% of L^*	0.0	0.0	21.4	74.9	2.9
% of U^*	0.0	0.0	3.6	27.3	68.3
% of patients	1.8	7.5	25.7	28.4	36.6

$\bar{n} = 61.8$; %good = 99.2%; %none = 0.8; %tox = 9.4; %eff = 57.1

to recommend at least one BED correctly 99.2% of time with an average of 61.8 patients. Within all in-trial patients, 57.1% of them had efficacy outcomes (> 72% decline in urinary PGE-M) and only 9.4% of them experienced DLTs. About 91% of patients are allocated at BEDs (400, 600, or 800 mg).

As a further illustration, we include here the details of one simulated trial. In the first stage, patients are enrolled one by one and are assigned the lowest dose (200 mg). If no DLT occurred, the next patient (cohort) is assigned the next higher dose level. The dose level continues to increase until the occurrence of a DLT. Once a DLT is observed, the FLW algorithm [5] is adopted to find the MTD, which will be assigned to the next patient (cohort). In this simulated trial, there is no DLT in the first five patients when the dose level is gradually increased from 200 to 800 mg. After including the data (no DLTs) from these five patients, the FLW algorithm recommends 800 mg as the MTD, which is given to the next patient. Since sixth patient did not experience DLT either, the FLW algorithm continues to recommend 800 mg as the MTD and so the seventh patient is treated at 800 mg. The procedure is repeated until the minimal sample size of stage one, 10, is reached. After the first 10 patients, no DLT ever occurred, and so there is sufficient evidence to conclude that 800 mg is the MTD. Stage one is concluded and its statistics are reported in Table 7.

Table 7 Demo trial: patient allocations and responses of one simulated trial

Dose level	200 mg	300 mg	400 mg	600 mg	800 mg
Stage 1: finding MTD; one patient per cohort					
Sample size	1	1	1	1	6
No. of DLT	0	0	0	0	0
No. of efficacy	0	0	0	1	5
Stage 2: finding BEDs; several patients per cohort					
Sample size	0	0	17	17	17
No. of DLT	0	0	2	0	3
No. of efficacy	0	0	9	14	13

The trial now moves to the second stage to identify the BEDs. As described in Sect. 2.3, based on the prior information and data from stage one, the set of BEDs is found to be {600 mg, 800 mg} and thus the next cohort is treated at the BEDs and adjacently admissible doses: {400 mg, 600 mg, 800 mg}, with one patient per dose. After collecting the data from this cohort of 3 patients, the BED is updated and it remains the same. Therefore, the same doses (400, 600, and 800 mg) are assigned to the next cohort of three patients. This procedure is repeated until the minimal sample size of a trial, 60, is reached. The statistics of stage two are reported in Table 7. The BEDs turned out to be the same for all 17 cohorts in stage two in this particular simulated trial, but it is not the case in other simulated trials. After collecting data from 61 patients (10 from stage one and $17 \times 3 = 51$ from stage two), the evidence is sufficiently strong to conclude that 600 mg, 800 mg are the BEDs (see Sect. 2.3 for details). The trial is terminated before reaching the maximum sample size and recommends 600 mg and 800 mg for further phase II investigation.

5 Conclusion and Discussion

In this paper, we propose a curve-free Bayesian design for identifying the BEDs, assuming only a monotonic dose–toxicity relationship and a unimodal or plateau dose–efficacy relationship. We avoid the complexity of joint modelling of the toxicity–efficacy association by utilizing the marginal distribution approach. In addition to the identification of BEDs, a confirmation cohort is recruited to provide a better estimate of the efficacy/toxicity rates of the BEDs to facilitate the selection of a recommended dose to phase II trials.

Our design is motivated by the following considerations. First of all, parametric models are not necessarily better for small samples, which are typical in early phase clinical trials. Although parametric models provide increased power over nonparametric models, they are usually not robust in that small departures from an assumed model can have drastic effects on the performance of the related statistical inference procedures. On the other hand, nonparametric models, while robust, lack power in small samples to provide meaningful results. To achieve a balance between power and robustness, we adopt the curve-free approach: instead of assuming a parametric dose–response relationship, we assume that the dose–toxicity curve is monotonically increasing and the dose–efficacy curve is unimodal. Our simulation studies show that our design is robust under these weaker assumptions.

Secondly, if identifying the BEDs is of primary interest, it may not be necessary to model the association between toxicity and efficacy outcomes. With limited information on the agent under study in early phase clinical trials, the challenge of specifying a prior on the toxicity–efficacy association will likely discourage the use of Bayesian designs. Therefore, we adopt a simpler phase Ia/Ib approach that dispense with the modelling of the toxicity–efficacy association. The phase Ia stage of our design focuses on finding the MTD and toxicity is the primary outcome. Upon identifying the MTD, the phase Ib stage is initiated to search for the BEDs among doses not exceeding the MTD, and efficacy becomes the primary outcome. A disadvantage of our phase Ia/Ib approach is that, without

utilizing the association between the toxicity and efficacy outcomes, a larger sample size may be required. This can be mitigated by implementing an algorithm that is capable of locating the MTD efficiently and accurately in the phase Ia stage, as was demonstrated in our simulation studies.

Thirdly, when identifying the BEDs is the goal, grouping doses could be a good idea since pooling the samples should lead to increased precision of estimated efficacy rate. Based on these considerations, we built the skeleton of CFBD.

Finally, the careful selection of prior distribution and its parameters fill the muscles of the CFBD. The well-known beta priors are chosen for toxicity/efficacy rates so that people can easily understand the models. In addition, the model-free efficacy curve and less-informative priors can increase the robustness of CFBD. Errors in the initial priors can be fixed quickly. Our simulation studies show the robustness of the proposed CFBD.

Several designs have been proposed for seamless phase I/II trials. One difference between phase Ia/Ib and phase I/II trials is that phase I/II trials requires statistical inferences on the efficacy of the selected dose, whereas phase Ia/Ib trials only recommend a likely dose range that has initial signals of biological activities beyond a target level; that is, a phase Ia/Ib trial is not required to establish the efficacy of a selected dose. The proposed CFBD in this paper is for one-agent phase Ia/Ib trials. Yet it can be easily extended to two-agent combination trials, under the assumptions that the dose-toxicity surface is marginally monotonic non-decreasing and the dose–efficacy surface is unimodal. In recent years, several dose-finding algorithms have been proposed to identify the maximum tolerated combination; however, adopting an efficient dose-finding algorithm in the first stage, such as that proposed by Lee et al. [10], is crucial to the success of the extension. To identify the set of best efficacious combinations, the dose–efficacy surface can be modelled using the equivalent of a step function in two-dimensional space and a similar utility function.

Acknowledgements Ying Lu is partially supported by NIH Grant 4P30CA124435. The Authors would like to thank the reviewers for their constructive comments and Professor Jin Hua, South China Normal University, for his help.

References

1. Ahn, C., Kang, S.-H. and Xie, Y. (2007). Optimal biological dose for molecularly-targeted therapies. In: Wiley encyclopedia of clinical trials. Wiley, Hoboken
2. Brock K, Billingham L, Copland M, Siddique S, Sirovica M, Yap C (2007) Implementing the Eff-Tox dose-finding design in the Matchpoint trial. *BMC Med Res Methodol* 17:112
3. Cook N, Hansen AR, Siu LL, Razak AR (2015) Early phase clinical trials to identify optimal dose and safety. *Mol Oncol* 9:997–1007
4. Fan SK, Chaloner K (2004) Optimal designs and limiting optimal designs for a trinomial response. *J Stat Plan Inference* 126:347–360
5. Fan SK, Lu Y, Wang YG (2012) A simple Bayesian decision-theoretic design for dose finding trials. *Stat Med* 31(28):3719–3730

6. Friedman HS, Kokkinakis DM, Pluda J, Friedman AH, Cokgor I, Haglund MM, Bigner DD, Schold SC (1998) Phase I trial of O6-benzylguanine for patients undergoing surgery for malignant glioma. *J Clin Oncol* 16:3570–3575
7. Hansen AR, Cook N, Amir E, Siu LL, Razak AR (2017) Determinants of the recommended phase 2 dose of molecular targeted agents. *Cancer* 123:1409–1415
8. Ivanova A (2003) A new dose-finding design for bivariate outcomes. *Biometrics* 59:1001–1007
9. Krysan K, Reckamp KL, Dalwadi H, Sharma S, Dohadwala M, Dubinett SM (2005) PGE2 activates MAPK/Erik pathway signaling and cell proliferation in non-small cell lung cancer cells in an EGF receptor-independent manner. *Cancer Res* 65:6275–6281
10. Lee BL, Fan SK, Lu Y (2017) A curve-free Bayesian decision-theoretic design for two-agent phase I trials. *J Biopharm Stat* 27(1):34–43
11. Macdougall J (2006) Analysis of dose-response studies—Emax model. In: Ting N (ed) *Dose finding in drug development. Statistics for biology and Health*. Springer, New York
12. Morita S, Thall PF, Muller P (2008) Determining the effective sample size of a parametric prior. *Biometrics* 64(2):595–602
13. Reckamp KL, Krysan K, Morrow JD, Milne GL, Newman RA, Tucker C, Elashoff RM, Dubinett SM, Figlin RA (2006) A phase I trial to determine the optimal biological dose of celecoxib when combined with erlotinib in advanced non-small cell lung cancer. *Clin Cancer Res* 12:3381–3388
14. Sachs JR, Mayawala K, Gadamssetty S, Kang SP, de Alwis DP (2016) Optimal dosing for targeted therapies in oncology: drug development cases leading by example. *Clin Cancer Res* 22(6):1318–1324
15. Thall P (2010) Bayesian models and decision algorithms for complex early phase clinical trials. *Stat Sci* 25(2):227–244
16. Thall PF, Cook JD (2004) Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 60:684–692
17. Tourneau CL, Lee JJ, Siu LL (2009) Dose escalation methods in phase I cancer clinical trials. *J Natl Cancer Inst* 101(10):708–720
18. Zang Y, Lee JJ (2017) A robust two-stage design identifying the optimal biological dose for phase I/II clinical trials. *Stat Med* 36(1):27–42
19. Zhang W, Sargent DJ, Mandrekar S (2006) An adaptive dose-finding design incorporating both toxicity and efficacy. *Stat Med* 25:2365–2383