Published in final edited form as:

Stat Med. 2017 January 15; 36(1): 27–42. doi:10.1002/sim.7082.

A robust two-stage design identifying the optimal biological dose for phase I/II clinical trials

Yong Zang[1,*,†] and **J. Jack Lee**[2,*,‡]

[1]Department of Biostatistics, Indiana University, Indianapolis, Indiana 46202, U.S.A

[2]Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A

Abstract

We propose a robust two-stage design to identify the optimal biological dose for phase I/II clinical trials evaluating both toxicity and efficacy outcomes. In the first stage of dose finding, we use the Bayesian model averaging continual reassessment method to monitor the toxicity outcomes and adopt an isotonic regression method based on the efficacy outcomes to guide dose escalation. When the first stage ends, we use the Dirichlet-multinomial distribution to jointly model the toxicity and efficacy outcomes and pick the candidate doses based on a three-dimensional volume ratio. The selected candidate doses are then seamlessly advanced to the second stage for dose validation. Both toxicity and efficacy outcomes are continuously monitored so that any overly toxic and/or less efficacious dose can be dropped from the study as the trial continues. When the phase I/II trial ends, we select the optimal biological dose as the dose obtaining the minimal value of the volume ratio within the candidate set. An advantage of the proposed design is that it does not impose a monotonically increasing assumption on the shape of the dose-efficacy curve. We conduct extensive simulation studies to examine the operating characteristics of the proposed design. The simulation results show that the proposed design has desirable operating characteristics across different shapes of the underlying true dose-toxicity and dose-efficacy curves. The software to implement the proposed design is available upon request.

Keywords

dose-finding; optimal biological dose; Bayesian adaptive design; isotonic regression; Phase I/II design

1. Introduction

The primary goal of a phase I clinical trial for cytotoxic agents is typically to identify the maximum tolerated dose (MTD) based on the toxicity outcomes. Then, a phase II trial often follows to examine the efficacy of the drug at the identified MTD or the recommended dose

*Correspondence to: Yong Zang, Department of Biostatistics, Indiana University and J. Jack Lee, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center.
†zangyong2008@gmail.com
‡jjlee@mdanderson.org

level. Conventionally, phase I and phase II trials are carried out separately. However, for the purpose of streamlining the drug development process and reducing the associated time and cost, there is a growing trend to integrate phase I and phase II trials into one trial. Several seamless phase I/II trial designs that jointly model the efficacy and toxicity outcomes of cytotoxic agents are available in the literature [1, 2, 3, 4, 5, 6].

Traditional dose-finding designs for cytotoxic agents, including the commonly used 3+3 design [7] and continual reassessment method (CRM) [8], assume that both efficacy and toxicity outcomes increase monotonically with the dose. However, the recent development of novel molecularly targeted agents (MTAs) challenges the traditional paradigm of dose-finding designs because the assumption of a monotonically increasing dose-efficacy relationship may not hold for MTAs. Many of the MTAs block the division of cancer cells by identifying and attacking specific pathways involved in tumor growth. As a result, the toxicity of MTAs can be minimal within the therapeutic dose range, and the dose-efficacy curves of MTAs may not follow monotonic patterns [9, 10, 11, 12]. For example, Friedman et al.[13] conducted a phase I trial for patients undergoing craniotomy for malignant glioma who received an MTA, O^6-benzylguanine (O^6-BG). In that trial, the efficacy of the agent was measured by the activity of the target enzyme O^6-alkylguanine-DNA alkyltransferase (AGT). The goal of the trial was to find the dose required to deplete tumor AGT activity in five O^6-BG dose levels ranging from 40 mg/m^2 to 120 mg/m^2. The agent demonstrated minimal toxicity and a non-monotonic dose-response relationship. A total of 30 patients were enrolled in the trial and only one observation of toxicity was reported. At the end of the trial, dose level 4 at 100 mg/m^2 was selected as the most efficacious dose.

Hence, a more reasonable goal of dose-finding trial designs for MTAs is finding the optimal biological dose (OBD), which is defined as the dose that has a desirable efficacy performance while still safeguarding patients with an acceptable toxicity profile [3, 14]. The OBD of an MTA is not always the highest dose and may appear in the middle of the investigational dose range. In practice, the dose-efficacy curves for MTAs are often expected to be unimodal or to plateau within the therapeutic dose range. Several dose-finding clinical trial designs that identify the OBD for MTAs have been proposed. For example, Braun [15] developed the bivariate CRM (bCRM) model by extending the traditional CRM with a flexible bivariate distribution that jointly models both the toxicity and efficacy outcomes. Hunsberger et al. [16] proposed the slope-sign design to guide dose finding based on the sign of the estimated local dose-efficacy curve. Zhang et al. [17] proposed the trinomial CRM to find the OBD; this method was further extended by Mandrekar et al. [18] for drug combination trials. Recently, Zang et al. [19] proposed three adaptive dose-finding designs for trials that evaluate MTAs.

Due to the non-monotonic dose-efficacy curve for MTAs and relatively small sample size of phase I trials, the toxicity-efficacy profiles of the candidate doses identified from phase I trials often retain a high level of uncertainty. Hence, a phase II dose-validation trial should follow to further validate the response profiles of the candidate doses and select the optimal dose. However, in spite of the rich body of literature on phase I dose-finding trial designs for MTAs [15, 16, 17, 18, 19], limited research has been conducted on seamless phase I/II designs integrating both dose-finding and dose-validation schemes for MTAs. Recently,

Hoering et al. [14, 20] proposed an integrated phase I/II trial design to assess the toxicity and efficacy outcomes for targeted agents. They used a conventional dose-finding algorithm to identify the MTD in the first stage, followed by a dose-validation stage that examines the toxicity and efficacy profiles for the dose levels around the identified MTD. Simulation results show that this design performs well under the restrictive condition that the OBD is in the neighborhood of the MTD. However, if the OBD is far away from the MTD, the performance of this design remains unclear.

Our study is motivated by a phase I clinical trial conducted at the University of California at Los Angeles Medical Center [21]. The purpose of the trial is to find the OBD of the MTA celecoxib combined with erlotinib in patients diagnosed with advanced non-small cell lung cancer. Celecoxib is an inhibitor that targets the rate-limiting enzyme Cox-2, which can regulate cellular proliferation, migration and invasion. Twenty-two patients were enrolled and treated with celecoxib at dose levels ranging from 200 mg to 800 mg, combined with a fixed dose of erlotinib. Both toxicity and efficacy endpoints were measured during the trial. Toxicity was determined by the dose-limiting toxicity and efficacy was measured by the biological acticity of the urinary prostaglandin E-M (PGE-M). No patients experienced toxicity and the dosage of 600 mg of celecoxib was selected as the OBD because it was associated with the maximal decrease in PGE-M. As reported by the author, “a phase II trial of celecoxib at 600 mg bid and erlotinib versus erlotinib plus placebo is planned”[21], p. 3387.

In this article, motivated by the celecoxib trial, we propose a Bayesian two-stage seamless phase I/II design to identify the OBD by jointly monitoring the toxicity and efficacy outcomes. This design comprises a dose-finding stage and a dose-validation stage. In the first stage, we use the Bayesian model averaging continual reassessment method (BMA-CRM) [22] to monitor toxicity outcomes, and use an isotonic regression method for dose escalation based on efficacy outcomes. When the dose escalation ends, we employ a Dirichlet-multinomial distribution to jointly model the toxicity and efficacy outcomes and select the candidate set based on a toxicity-efficacy volume ratio. The candidate set is then advanced to the second stage of dose validation, where more patients are randomized to validate the candidate dose set and identify the OBD. We note that although the proposed design is inspired by a phase I trial evaluating MTAs, the application of the proposed design is not restricted to MTAs. Indeed, because we use the nonparametric isotonic regression method, which makes little assumption on the dose-efficacy curve, the proposed design is applicable to both MTAs and cytotoxic agents.

The remainder of the article is organized as follows. In Section 2, we propose the seamless phase I/II design. In Section 3, we present a simulation study to investigate the operating characteristics of the proposed design and compare it with other existing designs. We conclude with a discussion in Section 4.

2. Seamless phase I/II design

2.1. The dose-finding stage

The proposed seamless design starts with a dose-finding stage to find the candidate doses with acceptable toxicity-efficacy profiles. Specifically, we use the BMA-CRM ([22]) to monitor the toxicity outcomes. This method initializes multiple parallel CRM models and then averages the estimates of toxicity probabilities using the BMA method to enhance the robustness of the conventional CRM. Let $(d_1, \ldots, d_J)$ denote a set of J pre-specified increasing doses for the agent under investigation and define $\hat{\pi}_{jT}$ as the BMA estimate of the toxicity rate at dose d_j. Then, with the highest acceptable toxicity rate ϕ_T, we construct

$$\mathscr{A}_T^{(1)} = \{j : \hat{\pi}_{jT} \leq \phi_T\}$$

as the admissible set of doses based on toxicity in the first stage.

After establishing $\mathscr{A}_t^{(1)}$, the next step is to implement dose escalation within the admissible set. We propose an isotonic regression method for dose escalation based on the efficacy outcomes. The purpose of the dose escalation procedure is to accumulate information for identifying the OBD, treat patients with doses that achieve a high therapeutic effect, and safeguard patients at the same time. We consider a unimodal or plateaued dose-efficacy curve. We define p_{jE} as the efficacy rate at dose level j and denote K as the highest dose level within the admissible set $\mathscr{A}_t^{(1)}$. Our dose-escalation goal is to find the most efficacious dose with the admissible set $\mathscr{A}_t^{(1)}$, the dose level j^* such that

$$p_{1E} \leq \cdots \leq p_{j^*E} \geq \cdots \geq p_{KE}. \quad (1)$$

We use the isotonic regression method to find dose j^*. However, the original isotonic regression requires a pre-specified location of the mode j^*, which is unknown in our setting. To overcome this limitation, we enumerate all K possible locations of j^*. Then, at each specified location $j^* = l$ for $l = 1, \ldots, K$, we take the double-sided isotonic regression [19, 23] to fit the accumulated efficacy outcomes that satisfy the order constraint (1) and obtain the corresponding set of estimates $\hat{p}_{jE}^{(l)}$ for $j = 1, \ldots, K$. Then, we select j^* with the least goodness-of-fit mean squared error

$$j^* = \operatorname{argmin}_{l \in (1,\ldots,K)} \sum_{j=1}^{K} (\hat{p}_{jE}^{(l)} - \overline{p}_{jE})^2,$$

where $\bar{p}_{jE}$ is the proportion estimate of p_{jE}. With the identified j^* in hand, our dose-escalation procedure can be described as follows:

1. Treat the first cohort of patients at the lowest dose level or at the physician-specified dose.
2. At the current dose level j, based on the toxicity outcomes, using the BMA-CRM to identify the admissible set $\mathscr{A}_t^{(1)}$.
3. Identify the dose level j^* within the admissible set $\mathscr{A}_t^{(1)}$ as the dose with the highest therapeutic effect while still safeguarding patients.
4. If $j^* > j$, escalate the dose level to $j + 1$; if $j^* < j$, de-escalate the dose level to $j-1$. If $j^* = j$, identify j^h as the highest tried dose. Then, retain dose level j if $j^* < j^h$; otherwise, escalate to $j + 1$ to explore more dose levels.
5. Repeat steps 2 to 4 until the maximum sample size for the dose-finding stage is reached.

Note that the proposed dose-escalation approach models the toxicity and efficacy outcomes separately. Theoretically, these two endpoints can also be modeled jointly. However, considering that the number of patients treated at each dose is small at the beginning of the trial, empirical experience from recent studies indicates that the joint modeling approach adds computational complexity but does not improve the performance of the dose-finding study [24, 25]. However, when dose escalation ends, we have already accumulated certain information about both the toxicity and efficacy outcomes. Hence, to precisely evaluate the toxicity-efficacy profiles, a joint modeling approach is implemented hereafter to borrow strength across these two endpoints for the purpose of identifying the candidate set of doses and selecting the OBD.

We use the Dirichlet-multinomial distribution to jointly model the toxicity and efficacy outcomes [26]. In particular, let $T = 0, 1$ and $E = 0, 1$ denote the binary toxicity and efficacy outcomes. We define that among a total of n_j patients treated at dose level d_j, r_{jte} of them have experienced the event ($T = t, E = e$) with the associated probability $p_{jte} = \Pr(T = t, E = e|d_j)$ for $t, e = 0, 1$. Denoting $r_j = (r_{j00}, r_{j01}, r_{j10}, r_{j11})$ and $p_j = (p_{j00}, p_{j01}, p_{j10}, p_{j11})$, we jointly model the toxicity and efficacy outcomes using the Dirichlet-multinomial distribution as

$$\begin{aligned} r_j|p_j &\sim \text{multinomial}(n_j, p_j) \\ p_j &\sim \text{Dirichlet}(\alpha_j), \end{aligned} \quad (2)$$

where $\alpha_j = (\alpha_{j00}, \alpha_{j01}, \alpha_{j10}, \alpha_{j11})$ is the hyperparameter and represents the prior information at dose level j. Then, after setting $p_{jT} = p_{j10} + p_{j11}$ and $p_{jE} = p_{j01} + p_{j11}$ as the marginal toxicity and efficacy rates, the posterior distributions of p_j, p_{jT} and p_{jE} are

$$\begin{aligned} p_j|r_j &\sim \text{Dirichlet}(\alpha_j+r_j), \\ p_{jT}|r_j &\sim \text{beta}\left(\sum_{e=0}^{1}(\alpha_{j1e}+r_{j1e}), \sum_{e=0}^{1}(\alpha_{j0e}+r_{j0e})\right), \\ p_{jE}|r_j &\sim \text{beta}\left(\sum_{t=0}^{1}(\alpha_{jt1}+r_{jt1}), \sum_{t=0}^{1}(\alpha_{jt0}+r_{jt0})\right). \end{aligned} \quad (3)$$

The Dirichlet-multinomial model is used to estimate the joint toxicity-efficacy probabilities and build a measure to evaluate the OBD. Following Yin et al., [3], we use a three-dimensional toxicity-efficacy volume ratio as the tradeoff measure that jointly evaluates the toxicity and efficacy outcomes. Specifically, at dose level j, after denoting $\gamma_j = p_{j01}/(1-p_{jT})$ as the probability of efficacy conditional on no toxicity, we define the three-dimensional volume ratio as

$$\omega_j = \frac{p_{jT}(1-p_{jE})(1-\gamma_j)}{(1-p_{jT})p_{jE}\gamma_j}. \quad (4)$$

Based on this expression, ω_j decreases when either p_{jE} or γ_j increases. In addition, a decreased p_{jT} can result in a decreased ω_j. Therefore, a smaller ω is always preferred when identifying the OBD. By introducing γ_j, the volume ratio ω_j incorporates the correlation between toxicity and efficacy into the consideration. Figure (1) illustrates how the value of ω_j varies with the correlation between the efficacy and toxicity outcomes. We specify p_{jT} = 0.3, p_{jE} = 0.2, 0.15 and let the correlation increase from −0.25 to 0.25. As shown in Figure (1), the correlation between efficacy and toxicity can significantly affect the value of ω_j. In particular, the volume ratio increased when the correlation increased. In other words, given that there is no toxicity, a lower correlation is preferred rather than a higher correlation. That is because, as the correlation increases, the same p_{jE} can result in a higher p_{jT}, which indicates a lower γ_j and increases ω_j. Based on Figure (1), we claim that ω_j is an appropriate statistic for jointly measuring the OBD and a lower ω_j indicates a better dose level. Specifically, the dose that yields the minimum estimate of ω_j within the admissible set is declared as the OBD. Figure (2) expresses the contour plot of the volume ratio with different toxicity and efficacy probabilities. Figure (2) indicates that the volume ratio is a trade-off measure to establish a compromise between the efficacy and toxicity outcomes. For example, when the toxicity rate increases from 0.1 to 0.2 and further to 0.3, to keep the volume ratio at 1, the corresponding requirements for the efficacy rates are 0.26, 0.35 and 0.41, respectively, when the correlation coefficient is fixed at 0.1.

At the end of the dose-finding stage, we identify the admissible set $\mathscr{A}_t^{(1)}$. Then, for each dose level $j \in (1, \cdots, K)$ within the admissible set, we obtain the estimate of ω_j (denoted as $\hat{\omega}_j$). Given τ as the highest acceptable value of ω_j, we build the candidate set of doses as

$$\mathscr{B} = \{j \in (1, \cdots, K) : (\hat{\omega}_j \leq \tau) \cap (n_j > 0)\}.$$

By adding $n_j > 0$, we tighten the safety evaluation of the trial by requiring that all the doses in the candidate set $\mathcal{B}$ are tested by at least one cohort of patients during the dose-finding stage. We select $\tau = 8$, which is calculated based on equation (4) with a highest acceptable toxicity rate $p_{jT} = 0.3$, a lowest acceptable efficacy rate $p_{jE} = 0.2$ and a speculated correlation coefficient of 0.1. Alternatively, if the correlation between the efficacy and toxicity is unknown, then we can use the data accumulated at the end of the dose-finding stage to estimate. Simulation studies indicate that the proposed design is not sensitive to the way that the correlation coefficient is determined (results not shown). Also, if the objective of the phase I/II trial is to compare a control treatment to the identified OBD, then we need to add the control arm to the candidate set $\mathcal{B}$. When the dose-finding stage ends, we seamlessly advance $\mathcal{B}$ and trigger the dose-validation stage.

2.2. The dose-validation stage

Once the candidate set $\mathcal{B}$ has been constructed, we seamlessly trigger the second stage of the trial for dose validation. The purpose of the dose-validation stage is to further validate the toxicity-efficacy profile of the candidate set and finally identify the OBD. To allocate patients during the second stage, we can use either approach: adaptive randomization or equal randomization. Adaptive randomization can shift the allocation of patients to more efficacious dose levels. However, the sample size in the phase I/II study typically is not large, which can yield less than the desired precision in estimating the treatment effect. Given this reason, we equally randomize the patients enrolled in the dose-selection stage to the candidate set. However, we note that the proposed design is not restricted to equal randomization and can easily accommodate adaptive randomization.

We continuously monitor the toxicity and efficacy outcomes during the dose-validation stage to update the candidate set $\mathcal{B}$ by excluding any overly toxic or less efficacious dose. The same Dirichlet-multinomial distribution, (2) and (3), is used with n_j to denote the patient data accumulating from the dose-finding stage until the current cohort of patients in the dose-validation stage. Hence, under the Bayesian framework, the Dirichlet-multinomial distribution seamlessly utilizes the accruing data in a “learn-as-we-go” fashion and thereby provides more precise estimates than single phase I and phase II trials conducted separately. The dose-validation procedure is summarized as follows:

1. Equally randomize a cohort of patients to the candidate set $\mathcal{B}$.
2. Update the posterior estimates of $\hat{p}_{jT}$ and $\hat{\omega}_j$ within $\mathcal{B}$.
3. Drop any dose j with either $\hat{p}_{jT} > \phi_T$ or $\hat{\omega}_j > \tau$ from $\mathcal{B}$. If all the doses have been dropped from the trial, terminate the trial early.
4. Repeat steps 1 to 3 until the maximum sample size for the dose-validation stage is reached.

Then, at the end of the dose-validation stage, the dose level with the minimum $\hat{\omega}_j$ is selected as the OBD. As the proposed design targets the OBD, we refer to it as the OBD-based design.

In the dose-validation stage, we only drop doses from the trial due to toxicity and/or futility and do not terminate the trial for superiority. That is because when a dose is promising, for the purpose of enhancing the individual ethics of the trial, it is often preferred to enroll more patients into the trial. However, if desired, the proposed design can accommodate an early stopping rule for superiority. Specifically, let ρ be the criteria of the volume ratio for stopping early for superiority. Then, during the dose-validation stage, the trial is terminated early if any $\hat{\omega}_j < \rho$, and that dose level j is then declared as the OBD.

In addition, to handle the control arm in the dose-validation stage, we need to keep the candidate set $\mathcal{B}$ updated but not alter the control arm during the trial. We define $\delta > 0$ as the marginal meaningful difference for ω_j. Then, when the trial ends and the OBD is identified, we estimate the volume ratios for the OBD and the control treatment respectively (denoted as $\hat{\omega}_j^{\dagger}$ and $\hat{\omega}_0$). We conclude that the OBD is promising if $\hat{\omega}_j^{\dagger} < \hat{\omega}_0 - \delta$; otherwise, we conclude that the OBD is unpromising.

3. Simulation studies

3.1. Operating characteristics

We conducted comprehensive simulation studies to investigate the operating characteristics of the proposed OBD-based design under 6 scenarios with different toxicity and efficacy profiles. We compared the OBD-based design with an MTD-based design and the bCRM [15]. The MTD-based design mimics the seamless phase I/II design proposed by Hoering et al. [14, 20] and is a two-stage design that targets the MTD at the first stage. Specifically, the first stage of the MTD-based design uses the BMA-CRM to identify the MTD based solely on the toxicity outcomes. Then, at the second stage, the same dose-validation scheme is applied for 2 doses at and below the MTD to determine the OBD by jointly modeling and evaluating the toxicity and efficacy outcomes. The bCRM jointly models the toxicity-efficacy outcomes through a flexible bivariate binomial distribution and adaptively allocates patients to the identified OBD throughout the phase I/II trial.

The maximum sample size in the simulation was 30 for the first stage and 120 for the second stage. The sample size was calibrated to maintain certain requirements for the dose selection percentages. In particular, we required at most 5% probability of selecting the incorrect doses if none dose was promising, and at least 50% probability of selecting the true OBD. We are aware that the conventional phase I/II trial typically requires a higher MTD selection percentage than 50%. However, considering that finding the OBD is generally more difficult due to the non-monotonic dose-efficacy relationship [19], we believe that 50% is an appropriate choice for the minimal requirement for identifying the OBD. Patients were enrolled into the trial in cohorts of size 3. We used the global cross-ratio model [27] to simulate the association between the toxicity and efficacy outcomes with the ratio fixed at 1.5, which corresponds to a weak positive correlation coefficient of about 0.1 as suggested by Dienstmann et al.[28]. Notice that this cross-ratio model has been used in other phase I/II designs [3]. We fixed the ratio for the purpose of simplifying the simulation setting. However, we note that the proposed design can handle a varying ratio as well as no modification. We specified the highest acceptable toxicity rate as $\phi_T = 0.3$ and the highest

acceptable volume ratio as $\omega_j = 8$ throughout the simulation. We also set the prior parameter $a_j = (0.25, 0.25, 0.25, 0.25)$ for the Dirichlet distribution at any dose level j. In addition, we used 5 skeletons in the BMA-CRM, with the following values:

$$(p_{1T}, p_{2T}, p_{3T}, p_{4T}, p_{5T}) = \begin{cases} (0.01, 0.05, 0.09, 0.15, 0.20), \textit{skeleton 1}; \\ (0.20, 0.30, 0.40, 0.50, 0.60), \textit{skeleton 2}; \\ (0.10, 0.20, 0.30, 0.40, 0.50), \textit{skeleton 3}; \\ (0.02, 0.06, 0.12, 0.30, 0.50), \textit{skeleton 4}; \\ (0.10, 0.08, 0.15, 0.20, 0.30), \textit{skeleton 5}. \end{cases}$$

Table 1 summarizes the simulation results based on 5,000 replicates, including the dose selection percentage, percentage of patients treated at each dose level, percentage of patients experiencing toxicity and efficacy, and average sample size of the whole trial (dose-finding stage + dose-validation stage). In Table 1, in addition to the toxicity and efficacy rates, we list the volume ratios for all the dose levels. The OBD, which obtains the minimum volume ratio while still safeguarding the patients, is emphasized by a boldface font. Notice that there can be 2 OBDs under some scenarios because the difference between the minimum volume ratio and the secondary minimum is negligible (within 0.1) in those scenarios.

Scenario 1 represents the circumstance in which there is no OBD. The OBD-based, MTD-based and bCRM designs obtained 2.6%, 3.5% and 4.5% false positive rates, respectively. We also notice that the OBD-based design obtained the smallest average sample size of 30.4, suggesting that most of the trials were terminated at the end of the first stage.

Scenarios 2 and 3 simulated the unimodal dose-efficacy curve. All the doses are safe in scenario 2; whereas dose 5 is overly toxic in scenario 3. The OBD is located at doses 2 and 3 in scenario 2 and at doses 1 and 2 in scenario 3. The OBD-based design outperformed the other designs in these scenarios. For example, in scenario 2, the OBD-based design had a satisfactory OBD selection percentage of 83.9% and allocated 68.5% of the patients to the OBD levels. Contrarily, the MTD-based and bCRM designs performed poorly, with respective OBD selection percentages of 6.7% and 20.5%, and respectively allocated only 17.2% and 34.6% patients to the OBD levels. The OBD-based design enrolled as many as 114.8 patients into the trial on average, which was 43.7 more patients than in the MTD-based design and 28.9 more than in the bCRM design. Also, the OBD-based design reported the lowest toxicity rate of 7.9% and the highest efficacy rate of 28.4%. In other words, the OBD-based design obtained the best operating characteristics under that setting. The simulation results in scenario 3 were similar.

Scenarios 4 and 5 simulate the cases in which the efficacy rate initially increases with the dose level and then plateaus. The OBD is located at dose 2 for scenario 4 and at doses 2 and 3 for scenario 5. The OBD-based design performed best in scenario 4. It yielded the highest OBD selection percentage of 52.8%; whereas the MTD-based design and bCRM design respectively achieved selection percentages of only 9.2% and 14.5%. For the allocation of patients, the OBD-based design allocated 37.0% of the patients to the OBD, which was 27.3% higher than in the MTD-based design and 15.8% higher than in the bCRM design. In scenario 5, all the designs had approximately the same OBD selection percentage, and the

MTD-based design allocated around 15% more patients to the OBD than the other two designs. Scenario 6 mimics the traditional monotonically increasing dose-efficacy curve. According to the simulation results, the OBD-based design was still the best design in terms of OBD selection and patient allocation. In addition, the MTD-based design also obtained good performance in this scenario with an OBD selection percentage of 55.2% and allocating 63.6% of the patients to the OBD, which is only 3.7% less than the percentage of patients allocated to the OBD in the OBD-based design. A reasonable explanation for the plausible performances of the MTD-based design under the last two scenarios is that the MTD and the OBD are close to each other in these two scenarios. Specifically, dose 3 is not only the OBD but also the MTD under both scenario 5 and scenario 6. Consequently, the MTD-based design allocated more patients to dose level 3 and resulted in good performances in OBD (MTD) identification and patient allocation.

In summary, the OBD-based design is a robust design that performs well, regardless of the shape of the dose-efficacy curve. In contrast, the MTD-based and bCRM designs are sensitive to the shape of the dose-efficacy curve and the locations of the OBD and MTD. In particular, the MTD-based design obtains good performance when the OBD is in the neighborhood of the MTD, and the bCRM design performs well when a monotonically increasing dose-efficacy curve holds. If the dose-efficacy curve is unimodal and the OBD is not close to the MTD (e.g., scenarios 2 and 3), neither of the two alternative designs work well. As we typically know little about the dose-efficacy curve, we recommend the use of the OBD-based design in practice, especially for MTAs with possibly non-monotonic dose-efficacy curves.

We stated that the OBD-based design can incorporate a control arm. To investigate this setting, we also conducted simulation studies of the OBD-based design when incorporating a control arm. As the bCRM cannot accommodate a control arm, the comparison was restricted to the OBD-based and MTD-based designs. Table 2 summarizes the simulation results. The same scenarios used in Table 1 appear in Table 2, except that we added a control arm with a toxicity rate of 0.25 and an efficacy rate of 0.2 in each scenario of Table 2. The simulation results were similar to the results shown in Table 1. When there was no promising dose (scenario 1), both designs selected the control arm with overwhelming percentages (over 95%). When the MTD and the OBD were different (scenarios 2, 3 and 4), the OBD-based design outperformed the MTD-based design. Otherwise, these two designs were comparable. In general, we still recommend the OBD-based design when a control arm is added.

3.2. Sensitivity analysis

We conducted a sensitivity analysis to investigate the performances of the OBD-based design with different criteria of early stopping for superiority, ρ and hyperparameter a_j. Notice that all the simulated trials in Tables 1 and 2 did not stop for superiority (e.g., $\rho = 0$). However, as we mentioned earlier, by adding a positive value for ρ, the trial can stop early for superiority to reduce the sample size if any volume ratio $\omega_j < \rho$. Hence, it is of interest to study the performance of the OBD-based design under various values of ρ other than 0.

Table 3 and Table 4 summarize the simulation results, with the former studying the original OBD-based design and the latter investigating the OBD-based design that incorporates a control arm. The results from these two tables are similar, so we focus on Table 3 hereafter. We take scenario 2 as an example. When there was no early stopping rule for superiority (ρ = 0), according to Table 1, the OBD-selection percentage was 89.3%, and 68.5% of the patients were allocated to the OBD levels. The average sample size was 114.8. This number dropped substantially to 49.0 when ρ = 0.25, according to the results in Table 3. Also, the OBD selection percentage and the proportion of patients treated at the OBD level decreased to 77.6% and 62.9% given ρ = 0.25. When ρ further increased from 0.25 to 1, the operating characteristics changed slightly, except that the average sample size decreased from 49.0 to 33.5. This result shows a trade-off for adding the early stopping rule for superiority. On one hand, adding this rule can result in substantial saving related to the number of patients enrolled in the trial and the associated resources. On the other hand, the OBD selection percentage and the proportion of patients treated at the OBD decrease when this rule is added.

We use the Dirichlet-multinomial distribution to jointly model the toxicity-efficacy outcomes. We adopt the non-informative prior by selecting α_j = (0.25, 0.25, 0.25, 0.25) for each dose level j. Conventionally, the non-informative prior is used when we lack information about a drug in advance of the trial. However, as a Bayesian design, it is also of interest to investigate the performance of the trial design with different prior information. To simplify the presentation, we use the prior parameters $p_T^{\text{pri}}=(p_{jT}^{\text{pri}}, j=1,\cdots,J), p_E^{\text{pri}}=(p_{jE}^{\text{pri}}, j=1,\cdots,J)$ and n^{pri} to represent the hyperparameter α_j. The values of p_{jT}^{pri} and p_{jE}^{pri} can be viewed as the initial guesses of the toxicity and efficacy response rates at dose j, and n^{pri} can be viewed as the number of patients treated at each dose level before the trial begins. Therefore, at dose level j, given $p_{jT}^{\text{pri}}, p_{jE}^{\text{pri}}$ and n^{pri} and assuming that the toxicity and efficacy outcomes are initially independent, the hyperparameter α_j can be represented as

$$\alpha_j=n^{\text{pri}}\left((1-p_{jT}^{\text{pri}})(1-p_{jE}^{\text{pri}}), (1-p_{jT}^{\text{pri}})p_{jE}^{\text{pri}}, p_{jT}^{\text{pri}}(1-p_{jE}^{\text{pri}}), p_{jT}^{\text{pri}}p_{jE}^{\text{pri}}\right).$$

Hence, as n^{pri} increases, the prior distribution becomes more informative. Notice that the conventional prior α_j = (0.25, 0.25, 0.25, 0.25) is equal to $p_{jT}^{\text{pri}}=p_{jE}^{\text{pri}}=0.5$ and n^{pri} = 1, which corresponds to a non-informative prior.

We examined the operating characteristics of the OBD-based design with different prior distributions. In particular, we fixed $p_T^{\text{pri}}=(0.1, 0.2, 0.3, 0.4, 0.5)$ and varied the values of p_E^{pri} and n^{pri}. Table 5 summarizes these results. We adopted scenario 2 from Table 1 to simulate the trial. Although the true dose-efficacy shape is unimodal, to better investigate the prior sensitivity, we considered a wide range of the prior dose-efficacy shapes such as monotonically increasing, unimodal and plateaued. Also, we considered different prior sample sizes with n_{pri} varying from 0.5 to 5. Therefore, n_{pri} = 0.5 corresponds to an

extremely non-informative prior and $n_{pri} = 5$ corresponds to a relatively informative prior. Based on Table 5, we found the results to be rather stable across different priors, suggesting that the OBD-based design is not sensitive to the specification of the prior.

4. Discussion

In this article, we propose a two-stage seamless design for phase I/II clinical trials in which we jointly model the toxicity and efficacy outcomes and use a toxicity-efficacy trade-off measure to identify the optimal biological dose. An important advantage of the proposed design is that it imposes little shape assumption on the dose-efficacy curve. Consequently, the proposed design is robust and yields plausible performances across different shapes of the underlying true dose-efficacy curves. In addition, the proposed design is flexible in the sense that it can easily incorporate a control arm. The simulation results show that the proposed design has good performances for identifying the optimal biological dose. The *R* code to implement the proposed OBD-based design is available by contacting the first author of this article.

We select $n_1 = 30$ and $n_2 = 120$ as the maximum sample sizes for the first and second stages. These sample sizes were determined through calibration studies to maintain the desirable operating characteristics of the proposed design such as having a high chances of sending promising doses to the second stage and identifying the optimal biological dose at the end of the trial. We notice that this sample size is generally larger than the total sample size of two separate phase I and phase II trials. However, according to the simulation results in Table 1, when there is no promising dose, most trials that use the proposed design will terminate at the end of the first stage, which will result in a reduced sample size. On the other hand, when the optimal biological dose exists, we believe that a slightly larger trial can still be desirable because more patients can benefit within the trial. Nevertheless, if a small-scale trial is preferred, the proposed design can be easily extended by adding an early stopping rule for superiority. According to the simulations, the maximum sample size for the second stage can be substantially reduced by adding an early stopping rule.

The proposed design is appropriate for trials in which the toxicity and efficacy outcomes are both binary. It is worth studying how to extend the proposed designs to handle other types of response outcomes, such as multiple-grade toxicity or time to disease progression. We focus on phase I/II trials that evaluate a single agent. It is also of interest to extend the proposed design to drug combination trials. In this article, we fix the sample size at 30 for the dose-finding stage; however, the proposed design can use various methods to adaptively alter the sample size of the first stage. One practical option is to specify a cut-off value such as $\phi = 0.8$ for the candidate dose set $\mathcal{B}$. Next, we need to update $\mathcal{B}$ after each step of the dose-escalation procedure. Then, if for every dose level j in $\mathcal{B}$ we have the posterior probability $P(\omega_j \leq \tau | \mathcal{D}) > \phi$, we can terminate the dose-finding stage early and forward $\mathcal{B}$ to the dose-validation stage. In contrast, if the maximum sample size is reached but there are still certain dose levels in $\mathcal{B}$ that do not meet the above condition, we can expand the cohorts for the dose-finding stage. This adaptive scheme enhances the flexibility of the phase I/II trial but also increases the computational burden. Further research in this area is warranted.

Acknowledgments

The authors thank two referees and the associate editor for their valuable comments, and LeeAnn Chastain for her editorial assistance. J. Jack Lee was supported in part by grant 016672 from the National Cancer Institute.

References

1. O'Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. Biometrics. 2001; 57:1018–1029. [PubMed: 11764240]
2. Thall PF, Cook J. Dose-finding based on toxicity-efficacy trade-offs. Biometrics. 2004; 60:684–693. [PubMed: 15339291]
3. Yin G, Li Y, Ji Y. Bayesian dose-finding in phase I/II trials using toxicity and efficacy odds ratio. Biometrics. 2006; 62:777–784. [PubMed: 16984320]
4. Huang X, Biswas S, Oki Y, Issa JP, Berry DA. A parallel phase I/II clinical trial design for combination therapies. Biometrics. 2007; 63:429–436. [PubMed: 17688495]
5. Yuan Y, Yin G. Bayesian phase I/II adaptively randomized oncology trials with combined drugs. Annals of Applied Statistics. 2011; 5:924–942. [PubMed: 22375162]
6. Pan H, Huang P, Wang Z, Wang L, Li C, Xia J. A novel Bayesian seamless phase I/II design. PLoS One. 2013; 8(9)
7. Storer BE. Design and analysis of phase I clinical trials. Biometrics. 1989; 45:925–937. [PubMed: 2790129]
8. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics. 1990; 46:33–48. [PubMed: 2350571]
9. Korn EL. Nontoxicity endpoints in phase I trial designs for targeted, non-cytotoxic agents. Journal of the National Cancer Institute. 2004; 96:977–978. [PubMed: 15240771]
10. Parulekar WR, Eisenhauer EA. Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: Theory and practice. Journal of the National Cancer Institute. 2004; 96:990–997. [PubMed: 15240782]
11. Tourneau CL, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. Journal of the National Cancer Institute. 2009; 101:708–720. [PubMed: 19436029]
12. LoRusso PM, Boerner SA, Seymour L. An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. Clinical Cancer Research. 2010; 16:1710–1718. [PubMed: 20215546]
13. Friedman HS, Kokkinakis DM, Pluda J, Friedman AH, Cokgor I, Haglund MM, Ashley DM, Rich J, Dolan EM, Pegg AE, Moschel RC, Mclendon RE, Kerby T, Herndon JE, Bgner DD, Schold SC. Phase I trial of O6-benzlguanine for patients undergoing surgery for malignant glioma. Journal of Clinical Oncology. 1998; 16:3570–2575. [PubMed: 9817277]
14. Hoering A, LeBlanc M, Crowley J. Seamless phase I–II trial design for assessing toxicity and efficacy for targeted agents. Clinical Cancer Research. 2011; 17:640–646. [PubMed: 21135145]
15. Braun TM. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. Controlled Clinical Trials. 2002; 23:240–256. [PubMed: 12057877]
16. Hunsberger S, Rubnstein LV, Dancey J, Korn EL. Dose escalation trial designs based on a molecularly targeted endpoint. Statistics in Medicine. 2005; 24:2171–2181. [PubMed: 15909289]
17. Zhang W, Sargent DJ, Mandrekar SJ. An adaptive dose-finding design incorporating both toxicity and efficacy. Statistics in Medicine. 2006; 25:2365–2383. [PubMed: 16220478]
18. Mandrekar SJ, Cui Y, Sargent DJ. An adaptive phase I design for identifying a biologically optimal dose for dual agent drug combinations. Statistics in Medicine. 2007; 26:2317–2330. [PubMed: 17016867]
19. Zang Y, Lee JJ, Yuan Y. Adaptive designs for identifying optimal biological dose for molecularly targeted agents. Clinical Trials. 2014; 11:319–327. [PubMed: 24844841]
20. Hoering A, Mitchell A, LeBlanc M, Crowley J. Early phase trial design for assessing several dose levels for toxicity and efficacy for targeted agents. Clinical Trials. 2013; 10:422–429. [PubMed: 23529697]

21. Reckamp KL, et al. A phase I trial to determine the optimal biological dose of celecoxib when combined with erlotinib in advanced non-small cell lung cancer. Clinical Cancer Research. 2006; 12:3381–3388. [PubMed: 16740761]
22. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. Journal of the American Statistical Association. 2009; 104:954–968.
23. Turner TR, Wollan PC. Locating a maximum using isotonic regression. Computational Statistics and Data Analysis. 1997; 25:305–320.
24. Cunana K, Koopmeiners JK. Evaluating the performance of copula models in phase I–II clinical trials under model misspecification. BMC Medical Research Methodology. 2014:51. [PubMed: 24731155]
25. Liu S, Johnson VE. A robust Bayesian dose-finding design for phase I/II clinical trials. Biostatistics. 2015 in press.
26. Thall PF, Simon R, Estey EH. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. Statistics in Medicine. 1995; 14:357–379. [PubMed: 7746977]
27. Dale JR. Global cross-ratio models for bivariate, discrete, ordered responses. Biometrics. 1986; 42:909–917. [PubMed: 3814731]
28. Dienstmann R, Brana I, Rodon J, Tabernero J. Toxicity as a biomarker of efficacy of molecular targeted therapies: focus on EGFR and VEGF inhibiting anticancer drugs. Oncologist. 2011; 16:1729–1740. [PubMed: 22135123]

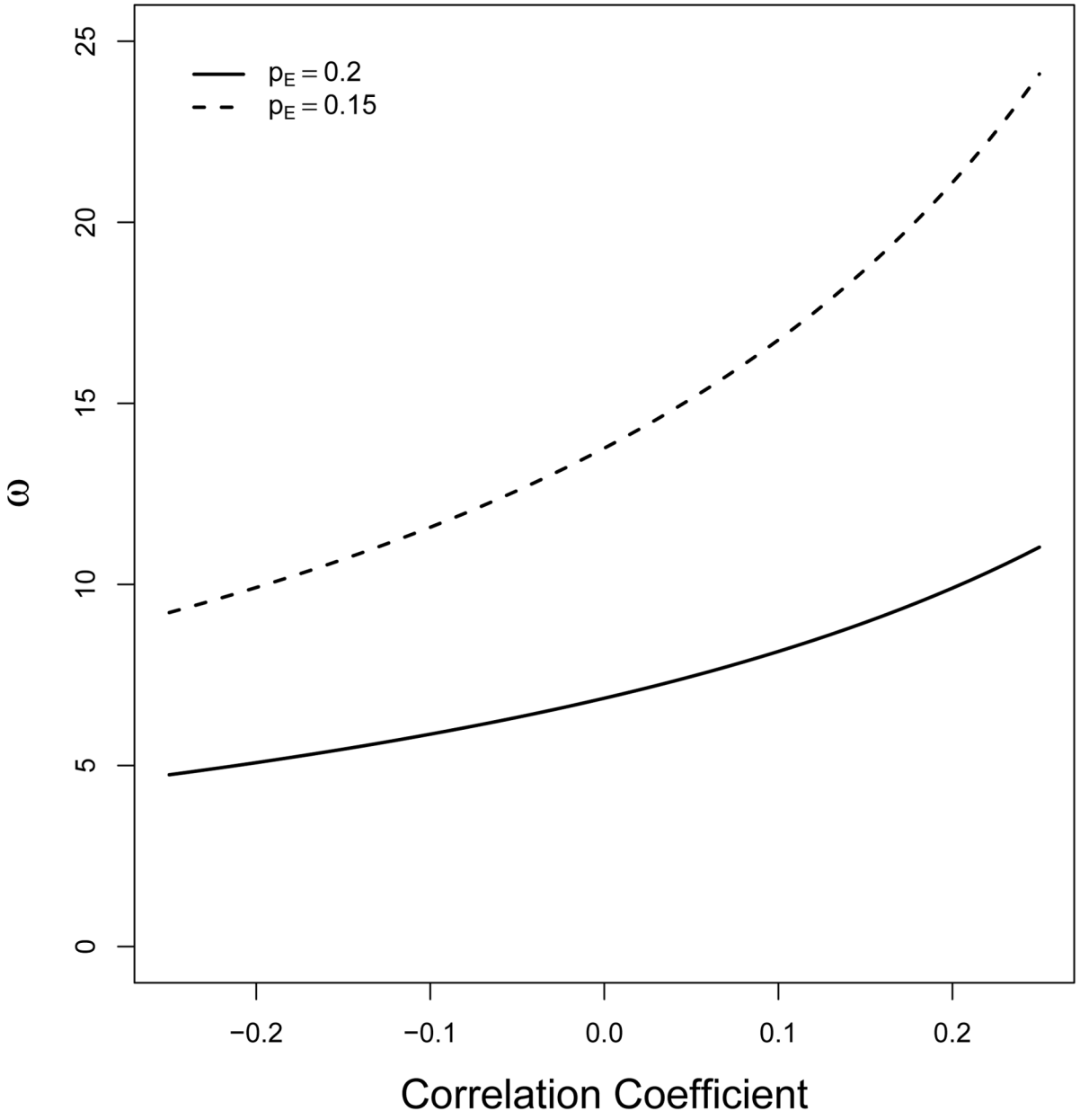

Figure 1.
Volume ratio with different correlation coefficient between efficacy and toxicity outcomes. The toxicity rate is fixed at $p_T = 0.3$.

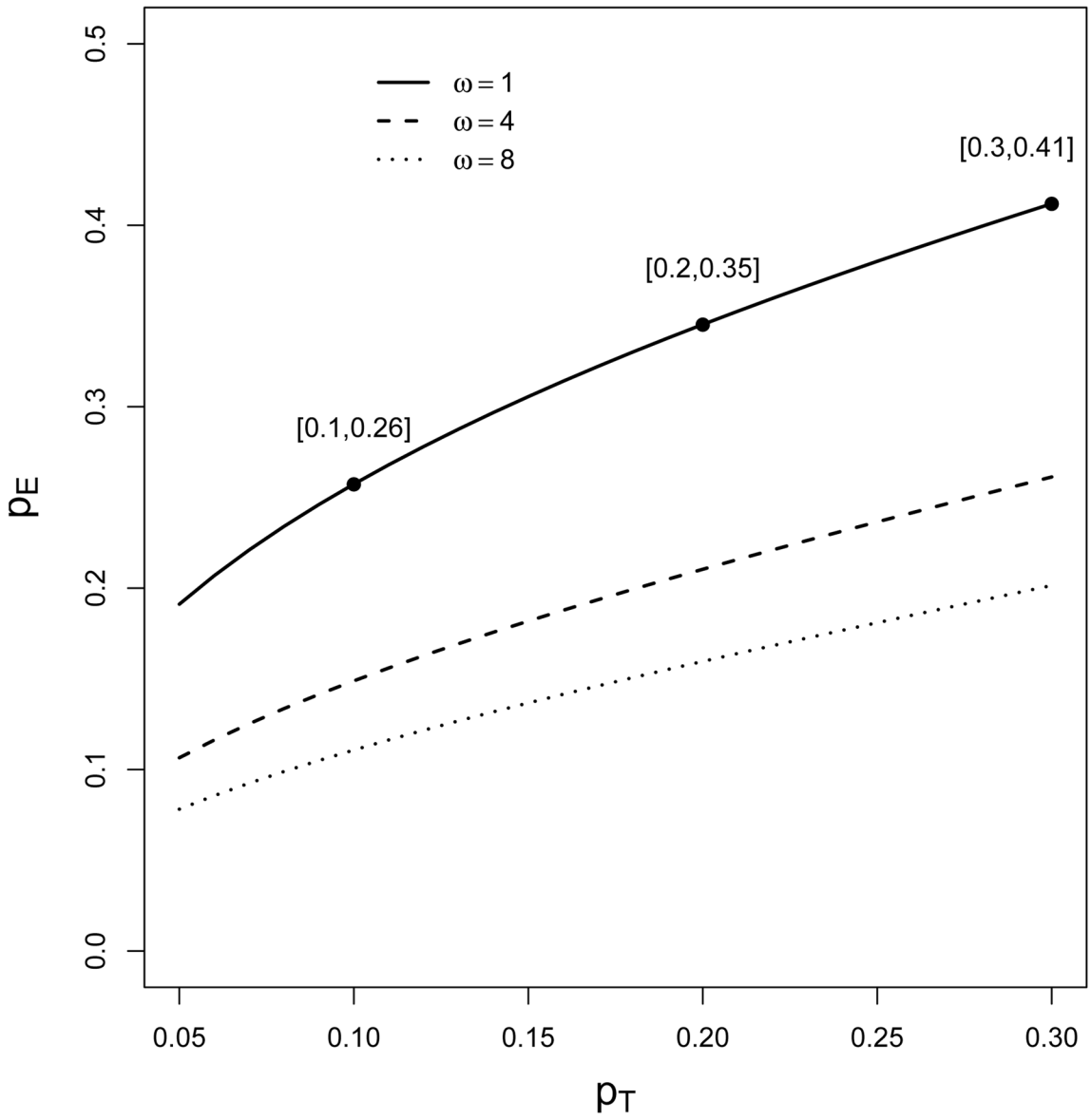

Figure 2.
Contour plot of the volume ratio ω with different toxicity and efficacy probabilities. The correlation coefficient is fixed at 0.1.

Table 1

The dose selection percentage, average percentage of patients treated at each dose level, percentages of toxicity and efficacy and average sample size under the OBD-based, MTD-based and bCRM designs.

		Dose level							
Design		**1**	**2**	**3**	**4**	**5**	**% of toxicity**	**% of efficacy**	**Average sample size**
					Scenario 1				
	True toxicity	0.1	0.2	0.3	0.4	0.5			
	True efficacy	0.05	0.1	0.18	0.25	0.3			
	Volume ratio	42.1	22.2	10.2	7.1	6.7			
OBD-based	Patients(%)	46.3	27.4	19.5	5.7	1.1	26.3	9.9	30.4
	Selection(%)	0.2	0.4	2.0	0	0			
MTD-based	Patients(%)	17.7	28.8	33.4	16.3	3.8	26.6	14.4	44.3
	Selection(%)	0.9	0.7	1.8	0.1	0			
bCRM	Patients(%)	34.6	35.4	20.7	7.3	2.0	27.6	10.4	33.9
	Selection(%)	0	0.8	3.6	0.1	0			
					Scenario 2				
	True toxicity	0.01	0.05	0.09	0.15	0.2			
	True efficacy	0.1	0.3	0.4	0.2	0.05			
	Volume ratio	0.8	**0.3**	**0.2**	3.0	99.1			
OBD-based	Patients(%)	17.1	32.2	36.3	12.0	2.4	7.9	28.4	114.8
	Selection(%)	7.9	**35.6**	**48.3**	2.8	0			
MTD-based	Patients(%)	6.0	6.4	10.8	41.8	35.0	14.9	16.0	71.1
	Selection(%)	0	**0.5**	**6.2**	32.8	0			
bCRM	Patients(%)	8.9	27.0	7.6	13.0	43.5	14.1	15.8	85.9
	Selection(%)	0.9	**19.6**	**0.9**	6.5	19.5			
					Scenario 3				
	True toxicity	0.02	0.06	0.12	0.3	0.5			
	True efficacy	0.3	0.4	0.2	0.1	0.05			
	Volume ratio	**0.1**	**0.1**	2.3	39.8	450.3			
OBD-based	Patients(%)	38.6	41.9	15.5	2.9	1.1	8.2	31.5	110.0
	Selection(%)	**41.0**	**46.2**	3.7	0	0			

Design		Dose level 1	2	3	4	5	% of toxicity	% of efficacy	Average sample size
MTD-based	Patients(%)	6.3	12.6	38.2	27.3	15.6	20.4	17.6	72.6
	Selection(%)	**0.1**	**11.0**	31.2	0.1	0			
bCRM	Patients(%)	12.3	37.2	13.9	29.8	6.8	18.4	23.5	81.0
	Selection(%)	**1.8**	**32.4**	1.7	10.7	0			
					Scenario 4				
	True toxicity	0.02	0.06	0.12	0.3	0.5			
	True efficacy	0.1	0.3	0.3	0.3	0.3			
	Volume ratio	1.7	**0.4**	0.8	2.7	6.7			
OBD-based	Patients(%)	19.5	37.0	32.5	9.3	1.7	11.9	28.0	108.6
	Selection(%)	7.3	**52.8**	27.2	1.0	0			
MTD-based	Patients(%)	4.4	9.7	43.6	31.3	11.0	20.7	28.5	95.1
	Selection(%)	0	**9.2**	48.3	9.1	0			
bCRM	Patients(%)	8.6	21.2	17.2	47.6	5.4	22.0	27.8	107.3
	Selection(%)	0.2	**14.5**	11.5	59.3	0.4			
					Scenario 5				
	True toxicity	0.1	0.2	0.25	0.4	0.5			
	True efficacy	0.2	0.4	0.4	0.4	0.4			
	Volume ratio	1.9	**0.7**	**0.8**	1.8	2.8			
OBD-based	Patients(%)	41.8	33.8	22.4	1.8	0.2	23.8	29.8	80.6
	Selection(%)	18.7	**41.0**	**15.8**	0	0			
MTD-based	Patients(%)	15.3	33.6	36.0	12.0	3.1	24.1	36.6	95.8
	Selection(%)	7.7	**33.4**	**25.7**	0.3	0			
bCRM	Patients(%)	44.5	48.5	4.5	2.1	0.4	25.4	34.4	82.4
	Selection(%)	0.5	**55.1**	**1.1**	0.2	0			
					Scenario 6				
	True toxicity	0.05	0.1	0.2	0.35	0.5			
	True efficacy	0.05	0.2	0.3	0.4	0.5			
	Volume ratio	19.5	**1.8**	**1.7**	1.4	1.2			
OBD-based	Patients(%)	23.6	35.7	31.6	7.9	1.2	19.0	23.9	90.5
	Selection(%)	0.9	**33.3**	**29.3**	0.9	0			

Design		Dose level					% of toxicity	% of efficacy	Average sample size
		1	2	3	4	5			
MTD-based	Patients(%)	5.6	18.2	45.4	23.5	7.3	23.3	29.5	88.4
	Selection(%)	0	**14.8**	**40.4**	2.3	0			
bCRM	Patients(%)	18.0	23.7	34.6	21.7	2.0	23.7	24.8	95.0
	Selection(%)	0	**20.9**	**36.7**	18.4	0			

Table 2

The dose selection percentage, average percentage of patients treated at each dose level, percentages of toxicity and efficacy and average sample size under the OBD-based and MTD-based designs incorporating the control arm.

		Dose level								
Design		**1**	**2**	**3**	**4**	**5**	**control**	**% of toxicity**	**% of efficacy**	**Average sample size**
						Scenario 1				
	True toxicity	0.1	0.2	0.3	0.4	0.5	0.25			
	True efficacy	0.05	0.1	0.18	0.25	0.3	0.2			
	Volume ratio	42.1	22.2	10.2	7.1	6.7	6.0			
OBD-based	Patients(%)	46.6	26.5	17.7	5.2	0.8	3.2	26.5	10.5	31.4
	Selection(%)	0.2	0.6	1.0	0	0	98.2			
MTD-based	Patients(%)	15.0	24.6	30.6	16.0	3.6	10.2	27.0	15.2	49.5
	Selection(%)	0.3	0.8	2.4	0	0	96.5			
						Scenario 2				
	True toxicity	0.01	0.05	0.09	0.15	0.2	0.25			
	True efficacy	0.1	0.3	0.4	0.2	0.05	0.2			
	Volume ratio	0.8	**0.3**	**0.2**	3.0	99.1	6.0			
OBD-based	Patients(%)	18.3	32.6	33.2	10.4	2.1	3.4	8.3	27.4	114.7
	Selection(%)	9.0	**38.6**	**44.6**	2.3	0	5.5			
MTD-based	Patients(%)	5.7	5.9	10.1	37.0	32.9	8.4	15.8	16.1	73.6
	Selection(%)	0	**0.5**	**6.2**	30.4	0	62.9			
						Scenario 3				
	True toxicity	0.02	0.06	0.12	0.3	0.5	0.25			
	True efficacy	0.3	0.4	0.2	0.1	0.05	0.2			
	Volume ratio	**0.1**	**0.1**	2.3	39.8	450.3	6.0			
OBD-based	Patients(%)	37.5	41.2	14.3	3.1	0.8	3.1	9.1	33.3	108.8
	Selection(%)	**38.9**	**47.4**	3.2	0	0	10.5			
MTD-based	Patients(%)	5.9	10.9	36.1	25.8	13.3	8.0	21.1	17.5	75.2
	Selection(%)	**0**	**9.1**	30.9	0	0	60.0			
						Scenario 4				
	True toxicity	0.02	0.06	0.12	0.3	0.5	0.25			

Design		Dose level 1	2	3	4	5	control	% of toxicity	% of efficacy	Average sample size
	True efficacy	0.1	0.3	0.3	0.3	0.3	0.2			
	Volume ratio	1.7	**0.4**	0.8	2.7	6.7	6.0			
OBD-based	Patients(%)	20.5	36.2	29.6	8.8	1.5	3.4	11.8	25.2	109.9
	Selection(%)	9.1	**52.6**	25.3	1.3	0	11.7			
MTD-based	Patients(%)	4.4	8.0	41.1	29.9	10.2	6.4	21.2	27.8	94.8
	Selection(%)	0	**6.9**	49.5	7.8	0	35.8			
					Scenario 5					
	True toxicity	0.1	0.2	0.25	0.4	0.5	0.25			
	True efficacy	0.2	0.4	0.4	0.4	0.4	0.2			
	Volume ratio	1.9	**0.7**	**0.8**	1.8	2.8	6.0			
OBD-based	Patients(%)	53.5	29.7	12.0	1.7	0.2	3.0	24.4	28.7	78.9
	Selection(%)	14.7	**38.7**	**14.4**	0.2	0	32.0			
MTD-based	Patients(%)	14.9	26.9	35.8	12.4	3.4	6.6	24.3	35.5	96.3
	Selection(%)	8.7	**26.5**	**27.9**	0.5	0	36.4			
					Scenario 6					
	True toxicity	0.05	0.1	0.2	0.35	0.5	0.25			
	True efficacy	0.05	0.2	0.3	0.4	0.5	0.2			
	Volume ratio	19.5	**1.8**	**1.7**	1.4	1.2	6.0			
OBD-based	Patients(%)	24.3	32.9	29.3	7.9	1.4	4.2	19.6	20.5	86.4
	Selection(%)	1.9	**31.4**	**27.6**	1.1	0	38.0			
MTD-based	Patients(%)	5.7	15.9	41.4	23.7	7.1	6.2	23.7	29.5	88.4
	Selection(%)	0	**12.8**	**38.5**	3.1	0	45.6			

Table 3

Sensitivity analysis for the OBD-based design under different ρ, the criteria for early stopping for superiority.

		Dose level							
ρ		1	2	3	4	5	% of toxicity	% of efficacy	Average sample size
						Scenario 1			
	True toxicity	0.1	0.2	0.3	0.4	0.5			
	True efficacy	0.05	0.1	0.18	0.25	0.3			
	Volume ratio	42.1	22.2	10.2	7.1	6.7			
0.25	Patients(%)	46.4	29.3	18.6	4.9	0.8	26.4	10.1	30.4
	Selection(%)	0.1	0.9	2.7	0.1	0			
0.5	Patients(%)	46.9	29.2	18.5	4.8	0.6	26.4	10.4	29.6
	Selection(%)	0.2	1.4	3.5	0.2	0.1			
1	Patients(%)	48.6	28.0	17.8	5.0	0.6	26.5	10.2	27.1
	Selection(%)	0.9	2.2	4.5	1.1	0.2			
						Scenario 2			
	True toxicity	0.01	0.05	0.09	0.15	0.2			
	True efficacy	0.1	0.3	0.4	0.2	0.05			
	Volume ratio	0.8	**0.3**	**0.2**	3.0	99.1			
0.25	Patients(%)	20.9	27.3	35.6	12.4	3.8	8.0	28.2	49.0
	Selection(%)	7.8	**31.3**	**46.3**	7.2	1.2			
0.5	Patients(%)	20.5	28.0	34.5	13.1	3.9	8.1	28.0	38.4
	Selection(%)	8.1	**31.0**	**46.3**	8.5	1.1			
1	Patients(%)	20.9	26.8	35.2	13.2	3.9	8.7	27.2	33.5
	Selection(%)	7.1	**30.7**	**45.0**	7.2	1.0			
						Scenario 3			
	True toxicity	0.02	0.06	0.12	0.3	0.5			
	True efficacy	0.3	0.4	0.2	0.1	0.05			
	Volume ratio	**0.1**	**0.1**	2.3	39.8	450.3			
0.25	Patients(%)	34.4	43.3	15.8	5.1	1.4	8.8	32.6	38.4
	Selection(%)	**31.5**	**52.0**	6.4	1.7	0.4			
0.5	Patients(%)	35.1	42.3	15.8	5.2	1.6	9.5	32.2	33.8

		Dose level							
ρ		1	2	3	4	5	% of toxicity	% of efficacy	Average sample size
	Selection(%)	**31.7**	**50.0**	6.7	1.9	0			
1	Patients(%)	34.3	42.1	16.7	5.5	1.4	9.4	32.0	32.2
	Selection(%)	**31.7**	**49.2**	7.5	2.8	0.3			
						Scenario 4			
	True toxicity	0.02	0.06	0.12	0.3	0.5			
	True efficacy	0.1	0.3	0.3	0.3	0.3			
	Volume ratio	1.7	**0.4**	0.8	2.7	6.7			
0.25	Patients(%)	22.8	34.2	28.8	11.9	2.3	12.2	26.8	58.8
	Selection(%)	7.4	**41.8**	29.8	7.6	2.5			
0.5	Patients(%)	22.0	32.1	27.9	15.0	3.0	13.2	26.9	42.7
	Selection(%)	7.8	**40.4**	33.0	7.3	1.7			
1	Patients(%)	23.7	32.7	26.2	14.2	3.2	13.2	26.3	34.4
	Selection(%)	8.7	**40.8**	27.8	9.3	2.8			
						Scenario 5			
	True toxicity	0.1	0.2	0.25	0.4	0.5			
	True efficacy	0.2	0.4	0.4	0.4	0.4			
	Volume ratio	1.9	**0.7**	**0.8**	1.8	2.8			
0.25	Patients(%)	52.4	30.5	13.5	3.3	0.3	24.1	31.1	50.1
	Selection(%)	19.2	**29.6**	**11.1**	1.2	0			
0.5	Patients(%)	50.5	29.7	14.8	4.3	0.7	23.8	31.5	38.7
	Selection(%)	20.7	**27.5**	**11.2**	1.2	0.1			
1	Patients(%)	50.9	28.2	15.8	4.5	0.6	23.7	31.0	29.4
	Selection(%)	20.5	**27.6**	**9.6**	1.1	0.2			
						Scenario 6			
	True toxicity	0.05	0.1	0.2	0.35	0.5			
	True efficacy	0.05	0.2	0.3	0.4	0.5			
	Volume ratio	19.5	**1.8**	**1.7**	1.4	1.2			
0.25	Patients(%)	25.5	30.8	31.4	10.4	1.9	19.1	22.6	68.1
	Selection(%)	1.5	**28.1**	**29.9**	5.8	0.7			
0.5	Patients(%)	25.0	32.1	29.3	11.3	2.3	19.0	22.6	56.0

ρ		Dose level					% of toxicity	% of efficacy	Average sample size
		1	**2**	**3**	**4**	**5**			
1	Selection(%)	1.9	**27.7**	**28.8**	7.5	0.9	19.9	23.0	39.4
	Patients(%)	28.5	28.6	27.5	12.9	2.5			
	Selection(%)	2.7	**27.9**	**27.6**	8.5	1.1			

Table 4

Sensitivity analysis for the OBD-based design incorporating the control arm under different ρ.

		Dose level								
ρ		1	2	3	4	5	Control	% of toxicity	% of efficacy	Average sample size
							Scenario 1			
	True toxicity	0.1	0.2	0.3	0.4	0.5	0.25			
	True efficacy	0.05	0.1	0.18	0.25	0.3	0.2			
	Volume ratio	42.1	22.2	10.2	7.1	6.7	6.0			
0.25	Patients(%)	48.7	26.9	17.1	4.8	1.0	1.5	26.3	10.1	28.0
	Selection(%)	0.3	0.4	1.1	0.2	0.1	97.9			
0.5	Patients(%)	48.7	25.8	17.5	5.5	0.9	1.6	26.0	10.0	27.9
	Selection(%)	0.4	0.7	2.9	0.5	0.1	95.4			
1	Patients(%)	46.4	28.1	17.7	5.2	0.8	1.8	26.3	9.6	27.6
	Selection(%)	0.5	0.5	1.8	0.3	0	96.9			
							Scenario 2			
	True toxicity	0.01	0.05	0.09	0.15	0.2	0.25			
	True efficacy	0.1	0.3	0.4	0.2	0.05	0.2			
	Volume ratio	0.8	**0.3**	**0.2**	3.0	99.1	6.0			
0.25	Patients(%)	20.7	26.9	33.9	12.8	3.3	2.4	8.8	27.8	45.4
	Selection(%)	5.2	**30.6**	**40.7**	6.3	0.5	16.7			
0.5	Patients(%)	20.1	27.5	33.3	13.1	3.5	2.4	8.7	27.3	37.0
	Selection(%)	6.5	**29.9**	**40.0**	8.1	0.7	14.8			
1	Patients(%)	19.3	27.4	33.5	13.2	4.2	2.4	8.7	27.1	33.6
	Selection(%)	7.4	**28.7**	**39.9**	6.4	0.5	17.1			
							Scenario 3			
	True toxicity	0.02	0.06	0.12	0.3	0.5	0.25			
	True efficacy	0.3	0.4	0.2	0.1	0.05	0.2			
	Volume ratio	**0.1**	**0.1**	2.3	39.8	450.3	6.0			
0.25	Patients(%)	36.3	41.1	14.8	4.5	1.3	2.0	9.8	32.2	37.1
	Selection(%)	**30.5**	**44.2**	5.6	1.6	0.2	17.9			
0.5	Patients(%)	31.2	44.1	16.7	4.6	1.3	2.1	9.4	31.4	34.6

		Dose level								
ρ		1	2	3	4	5	Control	% of toxicity	% of efficacy	Average sample size
	Selection(%)	**23.7**	**50.6**	7.9	1.8	0.3	15.7			
1	Patients(%)	32.8	43.0	15.5	5.0	1.3	2.4	9.4	30.9	32.2
	Selection(%)	**27.0**	**47.7**	6.0	1.8	0.4	17.1			
							Scenario 4			
	True toxicity	0.02	0.06	0.12	0.3	0.5	0.25			
	True efficacy	0.1	0.3	0.3	0.3	0.3	0.2			
	Volume ratio	1.7	**0.4**	0.8	2.7	6.7	6.0			
0.25	Patients(%)	23.8	31.8	27.9	12.0	2.2	3.3	12.5	26.4	57.2
	Selection(%)	8.6	**38.4**	24.5	5.3	0.7	22.5			
0.5	Patients(%)	20.5	33.1	28.1	13.1	2.6	2.6	12.7	26.5	43.2
	Selection(%)	8.0	**37.7**	26.2	6.6	1.5	20.0			
1	Patients(%)	22.8	32.5	25.8	13.3	2.9	2.7	13.4	25.5	34.0
	Selection(%)	7.2	**38.0**	27.5	6.9	1.9	18.5			
							Scenario 5			
	True toxicity	0.1	0.2	0.25	0.4	0.5	0.25			
	True efficacy	0.2	0.4	0.4	0.4	0.4	0.2			
	Volume ratio	1.9	**0.7**	**0.8**	1.8	2.8	6.0			
0.25	Patients(%)	51.5	29.1	13.7	3.2	0.5	2.0	23.0	30.5	48.3
	Selection(%)	17.5	**29.7**	**12.0**	0.6	0.2	40.0			
0.5	Patients(%)	47.0	31.0	15.2	4.2	0.5	2.1	22.9	31.8	36.3
	Selection(%)	18.0	**29.5**	**10.8**	0.9	0.1	40.7			
1	Patients(%)	45.3	31.2	16.3	4.4	0.5	2.3	23.5	30.8	29.2
	Selection(%)	17.3	**29.8**	**11.8**	1.2	0.1	39.8			
							Scenario 6			
	True toxicity	0.05	0.1	0.2	0.35	0.5	0.25			
	True efficacy	0.05	0.2	0.3	0.4	0.5	0.2			
	Volume ratio	19.5	**1.8**	**1.7**	1.4	1.2	6.0			
0.25	Patients(%)	25.9	31.4	28.8	9.9	1.9	2.1	19.0	21.6	62.8
	Selection(%)	1.0	**25.3**	**24.6**	4.4	0.5	44.2			
0.5	Patients(%)	26.6	28.9	28.3	11.7	2.2	2.3	19.7	22.9	49.7

ρ		Dose level						% of toxicity	% of efficacy	Average sample size
		1	**2**	**3**	**4**	**5**	**Control**			
1	Selection(%)	1.7	**23.0**	**27.1**	6.2	1.0	41.0	19.6	21.3	37.2
	Patients(%)	29.4	26.6	26.2	12.2	2.9	2.7			
	Selection(%)	2.2	**22.2**	**25.5**	7.6	0.9	41.6			

Table 5

Sensitivity analysis for the OBD-based design under different prior distributions with $p_T^{\text{pri}}=(0.1, 0.2, 0.3, 0.4, 0.5)$.

	Dose level							
	1	**2**	**3**	**4**	**5**	**% of toxicity**	**% of efficacy**	**Average sample size**
True toxicity	0.01	0.05	0.09	0.15	0.2			
True efficacy	0.1	0.3	0.4	0.2	0.05			
Volume ratio	0.8	**0.3**	**0.2**	3.0	99.1			
n^{pri} = 0.5, $p_E^{\text{pri}}=(0.1, 0.2, 0.3, 0.3, 0.3)$								
Patients(%)	18.9	31.3	36.5	11.1	2.2	8.1	28.2	112.8
Selection(%)	16.7	32.3	41.4	2.6	0			
n^{pri} = 1, $p_E^{\text{pri}}=(0.1, 0.2, 0.3, 0.3, 0.3)$								
Patients(%)	18.9	32.0	36.9	9.8	2.4	7.8	28.0	113.7
Selection(%)	13.8	34.0	43.4	2.5	0			
n^{pri} = 2, $p_E^{\text{pri}}=(0.1, 0.2, 0.3, 0.3, 0.3)$								
Patients(%)	16.7	33.6	36.1	11.8	1.8	8.0	28.1	115.2
Selection(%)	10.5	37.6	44.3	2.8	0			
n^{pri} = 5, $p_E^{\text{pri}}=(0.1, 0.2, 0.3, 0.3, 0.3)$								
Patients(%)	14.1	34.7	41.1	8.8	1.3	7.9	30.2	116.8
Selection(%)	13.0	35.0	48.2	0.9	0			
n^{pri} = 0.5, $p_E^{\text{pri}}=(0.2, 0.3, 0.4, 0.2, 0.1)$								
Patients(%)	16.4	33.2	38.5	10.5	1.4	8.2	28.9	113.7
Selection(%)	13.6	34.4	43.6	2.3	0			
n^{pri} = 1, $p_E^{\text{pri}}=(0.2, 0.3, 0.4, 0.2, 0.1)$								
Patients(%)	17.4	33.0	38.1	10.1	1.4	8.0	29.0	114.3
Selection(%)	11.1	35.3	45.7	2.5	0			

	Dose level					% of toxicity	% of efficacy	Average sample size
	1	2	3	4	5			
$n^{\text{pri}} = 2$, $p_E^{\text{pri}} = (0.2, 0.3, 0.4, 0.2, 0.1)$								
Patients(%)	20.1	33.3	37.8	7.7	1.1	7.5	28.8	116.2
Selection(%)	10.3	36.7	46.3	0.2	0			
$n^{\text{pri}} = 5$, $p_E^{\text{pri}} = (0.2, 0.3, 0.4, 0.2, 0.1)$								
Patients(%)	28.4	31.5	32.9	6.5	0.7	7.1	26.7	115.4
Selection(%)	7.5	36.1	48.5	0	0			
$n^{\text{pri}} = 0.5$, $p_E^{\text{pri}} = (0.05, 0.1, 0.2, 0.25, 0.3)$								
Patients(%)	18.0	32.2	36.8	10.9	2.1	8.1	28.3	113.4
Selection(%)	15.6	36.0	40.1	1.9	0			
$n^{\text{pri}} = 1$, $p_E^{\text{pri}} = (0.05, 0.1, 0.2, 0.25, 0.3)$								
Patients(%)	17.7	32.9	36.3	11.1	2.0	8.2	28.2	113.7
Selection(%)	12.9	36.7	42.0	2.3	0			
$n^{\text{pri}} = 2$, $p_E^{\text{pri}} = (0.05, 0.1, 0.2, 0.25, 0.3)$								
Patients(%)	17.3	32.6	36.9	10.9	2.3	8.1	28.0	113.4
Selection(%)	10.7	35.9	44.4	2.3	0			
$n^{\text{pri}} = 5$, $p_E^{\text{pri}} = (0.05, 0.1, 0.2, 0.25, 0.3)$								
Patients(%)	16.8	33.2	39.1	8.8	2.1	8.0	28.6	114.1
Selection(%)	7.3	34.6	45.8	1.3	0			