

Bayesian Optimal Interval Designs for Phase I Clinical Trials

Suyu Liu and Ying Yuan

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

Summary. In phase I trials, effectively treating patients and minimizing the chance of exposing them to subtherapeutic and overly toxic doses are clinician's top priority. Motivated by this practical consideration, we propose Bayesian optimal interval (BOIN) designs to find the maximum tolerated dose (MTD) and minimize the probability of inappropriate dose assignments for patients. We show, both theoretically and numerically, that the BOIN design not only has superior finite- and large-sample properties, but also can be easily implemented in a simple way similar to the traditional "3+3" design. Compared to the well-known continual reassessment method, the BOIN design yields comparable average performance to select the MTD, but has a substantially lower risk of assigning patients to subtherapeutic and overly toxic doses. We apply the BOIN design to two cancer clinical trials.

Keywords: Decision error; Bayesian adaptive design; Dose finding; Maximum tolerated dose.

1. Introduction

Numerous phase I trial designs have been proposed to identify the maximum tolerated dose (MTD) of a new drug, which is typically defined as the dose with a dose-limiting toxicity probability that is closest to the target toxicity rate. Examples include the standard "3+3" design (Storer, 1989; Korn et al., 1994), the continual reassessment method (CRM; O'Quigley, Pepe and Fisher, 1990), the biased coin design (Durham and Flournoy, 1994), the decision-theoretic approach (Whitehead and Brunier, 1995), isotonic designs (Leung and Wang, 2001), and "A+B" designs (Lin and Shih, 2001), among others. Chevret (2006), Ting (2006) and Gerke and Siedentop (2008) provided comprehensive reviews of dose-finding methods for phase I clinical trials.

The interval design is a relatively new class of phase I trial designs, for which the dose transition is defined by the relative location of the observed toxicity rate (i.e., the number of patients who experienced toxicity divided by the total number of patients treated) at the current dose with respect to a prespecified toxicity tolerance interval (Oron, Azriel and Hoff, 2011). If the observed toxicity rate is located within the interval, we retain the current dose; if the observed toxicity rate is greater than the upper boundary of the interval, we deescalate the dose; and if the observed toxicity rate is smaller than the lower boundary of

Address for correspondence: Ying Yuan, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.
Email: yyuan@mdanderson.org

the interval, we escalate the dose. Yuan and Chappell (2004) suggested using a (tolerance) interval to determine dose escalation and deescalation. Ivanova, Flournoy and Chung (2007) developed a more formal interval design, called the cumulative cohort design. That design was further extended to ordinal and continuous outcomes based on a t -statistic (Ivanova and Kim, 2009). Neuenschwander, Branson and Gsponer (2008) proposed a practical and flexible Bayesian dose-finding method that divides the posterior distribution of the toxicity probability into four intervals (i.e., under-dosing, targeted toxicity, excessive toxicity and unacceptable toxicity) for decision making based on a two-parameter model. A similar posterior-interval approach was also adopted in the modified toxicity probability interval design (Ji et al., 2010).

One advantage of interval designs is that they are simple to implement in practice. Because the interval is prespecified, during trial conduct, the decision of which dose to administer to the next cohort of patients does not require complicated computations, but only a simple comparison of the observed toxicity rate at the current dose with the prespecified interval boundaries. Compared to other similarly easy-to-implement designs, e.g., the standard “3+3” design, the interval design is more efficient because it uses all the information from the cumulative data at the current dose to determine the next dose assignment. Recently, Oron, Azriel and Hoff (2011) showed that the interval design provides a convergence property similar to that of the CRM. The interval design converges almost surely, at a \sqrt{n} rate, to exclusive allocations at a dose level for which the true toxicity rate is within the interval. Under the interval design, the selection of an appropriate tolerance interval is critical for the performance of the design because the dose transition and assignment are entirely determined by this prespecified interval. Ivanova, Flournoy and Chung (2007) suggested selecting the interval by an exhaustive numerical search over the group up-and-down designs based on its large-sample property. However, sample sizes for phase I trials are small, typically less than 40.

In this article, we propose a flexible, *finite-sample-based* approach to selecting the interval boundaries, and use it to develop two Bayesian optimal interval (BOIN) designs. Our approach is motivated by top priority and concern of clinicians, which is to effectively treat patients and minimize the chance of exposing them to subtherapeutic and overly toxic doses.

The idea behind the BOIN designs is straightforward. The conduct of a phase I trial can be viewed as a sequence of decision-making steps of dose assignment for patients who are sequentially enrolled into the trial. At each moment of decision making, based on the observed data, we take one of three actions: escalate, deescalate or retain the current dose. Under the standard assumption that efficacy monotonically increases with toxicity, an ideal trial design would escalate the dose when the current dose is below the MTD in order to avoid treating a patient at subtherapeutic dose levels; deescalate the dose when the current dose is above the MTD in order to avoid exposing a patient to overly toxic doses; and retain the same dose level when the current dose is equal (or close) to the MTD. However, such an ideal design is not available in practice because we do not know whether the current dose is actually below, above or equal (or close) to the MTD, and need to infer that information and make decisions based on the data collected from patients who have been enrolled in the trial. Given the randomness of the observed data and small sample sizes of phase I trials, the decisions of dose assignment we make are often incorrect, e.g., we may escalate (or deescalate) the dose when it is actually higher (or lower) than the MTD, which results in overly aggressive (or conservative) dose escalation and treating excessive numbers of patients at dose levels above (or below) the MTD. From a practical and ethical viewpoint,

it is highly desirable to minimize such decision errors so that the actual design behaves as closely as possible to the ideal (error-free) design. The BOIN designs are developed to achieve this goal.

We consider two motivating cancer clinical trials. The first one is a phase I dose-finding clinical trial protocol recently initiated at MD Anderson Cancer Center. The goal of the trial is to identify the MTD of an aurora kinase inhibitor (AKI) as monotherapy, in combination with docetaxel in patients with advanced solid tumors, with a target toxicity rate of 25%. In this trial, docetaxel will be administered via intravenous infusion at a fixed dosage of 75mg/m². A total of six dose levels (i.e., 20, 40, 80, 150, 250, 350 mg) of AKI will be investigated and administered on days 1, 8 and 15 of a 28-day treatment cycle. Severity of adverse event or toxicity will be scored according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4.0. The dose-limiting toxicities (DLTs) include all grade 3 or 4 adverse events, grade ≥ 3 hematologic toxicities, grade ≥ 3 asthenia/fatigue, grade ≥ 3 proteinuria, and grade ≥ 2 neurotoxicity. A total of 36 patients will be treated in cohort sizes of 3.

The second motivating example is a pediatric phase I clinical trial (Jakacki et al., 2008), which aimed to determine the MTD of erlotinib for children with refractory solid tumors. Erlotinib is an oral inhibitor of the epidermal growth factor receptor signal pathway and has been approved by the Food and Drug Administration for adults with recurrent non-small-cell lung cancer and advanced pancreatic cancer. The trial studied five dose levels of erlotinib: 35, 50, 65, 85 and 110 mg. A total of 19 assessable patients were used for dose escalation. Determinations of DLTs included any grade 3 or 4 thrombocytopenia or grade 4 neutropenia, or any grade 3 or 4 nonhematologic toxicity. The target toxicity rate was 25%.

The remainder of this article is organized as follows. In Section 2, we formulate dose finding as a sequential decision-making problem and propose designs to minimize the decision errors of dose assignment during trial conduct. In Section 3, we present simulation studies to compare the operating characteristics of the new designs with those of some available designs, and apply the new method to two cancer clinical trials in Section 4. We conclude with a brief discussion in Section 5.

2. Method

2.1. Interval design

Assume that a total of J prespecified doses are under investigation, and let ϕ denote the target toxicity rate specified by physicians. We assume that patients are treated in cohorts but allow the cohort size to vary from one cohort to another. Interval designs can be generally described as follows:

- (a) Patients in the first cohort are treated at the lowest dose level.
- (b) At the current dose level j , assume that a total (or the cumulative number) of n_j patients have been treated, and m_j of them have experienced toxicity. Let $\hat{p}_j = m_j/n_j$ denote the observed toxicity rate at dose level j , and $\lambda_{1j}(n_j, \phi)$ and $\lambda_{2j}(n_j, \phi)$ denote the prespecified lower and upper (or dose escalation and deescalation) boundaries of the interval, respectively, with $0 \leq \lambda_{1j}(n_j, \phi) < \lambda_{2j}(n_j, \phi) \leq 1$. To assign a dose to the next cohort of patients,
 - if $\hat{p}_j \leq \lambda_{1j}(n_j, \phi)$, we escalate the dose level to $j + 1$;

- if $\hat{p}_j \geq \lambda_{2j}(n_j, \phi)$, we deescalate the dose level to $j - 1$;
- otherwise, i.e., $\lambda_{1j}(n_j, \phi) < \hat{p}_j < \lambda_{2j}(n_j, \phi)$, we retain the same dose level, j .

To ensure that the dose levels of the treatment always remain within the prespecified dose range, the dose escalation/deescalation rule requires some adjustments when j is at the lowest or highest level. That is, if $j = 1$ and $\hat{p}_j \geq \lambda_{2j}(n_j, \phi)$ or $j = J$ and $\hat{p}_j \leq \lambda_{1j}(n_j, \phi)$, the dose remains at the same level, j .

- (c) This process is continued until the maximum sample size is reached or the trial is terminated due to excessive toxicity, as described below.

This design is more general than the existing interval designs (e.g., Ivanova, Flournoy and Chung, 2007; Oron, Azriel and Hoff, 2011) because it allows the interval boundaries $\lambda_{1j}(n_j, \phi)$ and $\lambda_{2j}(n_j, \phi)$ to depend on both dose level j and n_j (i.e., the number of patients that have been treated at dose level j), whereas the existing interval designs assume that the interval boundaries are independent of j and n_j .

2.2. Local BOIN design

In the interval design, the selection of the interval boundaries $\lambda_{1j}(n_j, \phi)$ and $\lambda_{2j}(n_j, \phi)$ is critical because these two parameters essentially determine the operating characteristics of the design. To simplify the notations, we suppress the arguments n_j and ϕ in $\lambda_{1j}(n_j, \phi)$ and $\lambda_{2j}(n_j, \phi)$. In the following subsection, we describe a method to select λ_{1j} and λ_{2j} to minimize incorrect decisions of dose escalation and deescalation during the trial conduct. For convenience, we call the resulting design the local BOIN design because the optimization is anchored to three point (or local) hypotheses.

2.2.1. Minimizing the local decision error

In order to minimize incorrect decisions of dose assignment, we first formally define the correct and incorrect decisions. Letting p_j denote the true toxicity probability of dose level j for $j = 1, \dots, J$, we formulate three point hypotheses:

$$\begin{aligned} H_{0j} : & \quad p_j = \phi \\ H_{1j} : & \quad p_j = \phi_1 \\ H_{2j} : & \quad p_j = \phi_2, \end{aligned} \tag{1}$$

where ϕ_1 denotes the highest toxicity probability that is deemed subtherapeutic (i.e., below the MTD) such that dose escalation should be made, and ϕ_2 denotes the lowest toxicity probability that is deemed overly toxic such that dose deescalation is required. For simplicity, we will alter the notation slightly by dropping subscript j from H_{0j} , H_{1j} and H_{2j} wherever such alterations do not cause confusion.

Specifically, H_0 indicates that the current dose d_j is the MTD and we should retain the current dose to treat the next cohort of patients; H_1 indicates that the current dose is subtherapeutic (or below the MTD) and we should escalate the dose; and H_2 indicates that the current dose is overly toxic (or above the MTD) and we need to deescalate the dose. Therefore, the correct decisions under H_0 , H_1 and H_2 are retainment, escalation and deescalation (each based on the current dose level), denoted as \mathcal{R} , \mathcal{E} and \mathcal{D} , respectively. Correspondingly, the incorrect decisions under H_0 , H_1 and H_2 are $\bar{\mathcal{R}}$, $\bar{\mathcal{E}}$ and $\bar{\mathcal{D}}$, respectively,

where $\bar{\mathcal{R}}$ denotes the decisions complementary to \mathcal{R} (i.e., $\bar{\mathcal{R}}$ includes \mathcal{E} and \mathcal{D}), and $\bar{\mathcal{D}}$ and $\bar{\mathcal{R}}$ denote the decisions complementary to \mathcal{D} and \mathcal{R} , respectively.

We note that the purpose herein of specifying three hypotheses, H_0, H_1 and H_2 , is not to represent the truth and conduct hypothesis testing, but just to indicate the cases of special interest under which we optimize the performance of our design. In particular, H_1 and H_2 , or more precisely $\delta_1 = \phi - \phi_1$ and $\delta_2 = \phi_2 - \phi$, represent the minimal differences (or effect sizes) of practical interest to be distinguished from the target toxicity rate ϕ (or H_0), under which we want to minimize the average decision error rate for the trial conduct. A difference smaller than δ_1 and δ_2 may not be of practical importance, and a larger difference will lead to a smaller error rate because it is more easily distinguished from ϕ than ϕ_1 and ϕ_2 . This approach is analogous to sample size determination, for which we first specify a point alternative hypothesis to represent the minimal effect size of interest and then determine the sample size to ensure a desirable power under that hypothesis.

Under the Bayesian paradigm, we assign each of the hypotheses a prior probability of being true, denoted as $\pi_{kj} = \text{pr}(H_{kj}), k = 0, 1, 2$. Then under the dose assignment rule described in Section 2.1, the probability of making an incorrect decision (the decision error rate), denoted as $\alpha(\lambda_{1j}, \lambda_{2j})$, at each of the dose assignments is given by

$$\begin{aligned} \alpha(\lambda_{1j}, \lambda_{2j}) &= \text{pr}(H_{0j})\text{pr}(\bar{\mathcal{R}}|H_{0j}) + \text{pr}(H_{1j})\text{pr}(\bar{\mathcal{E}}|H_{1j}) + \text{pr}(H_{2j})\text{pr}(\bar{\mathcal{D}}|H_{2j}) \\ &= \text{pr}(H_{0j})\text{pr}(m_j \leq n_j\lambda_{1j} \text{ or } m_j \geq n_j\lambda_{2j}|H_{0j}) + \text{pr}(H_{1j})\text{pr}(m_j > n_j\lambda_{1j}|H_{1j}) \\ &\quad + \text{pr}(H_{2j})\text{pr}(m_j < n_j\lambda_{2j}|H_{2j}) \\ &= \pi_{0j}\{Bin(n_j\lambda_{1j}; n_j, \phi) + 1 - Bin(n_j\lambda_{2j} - 1; n_j, \phi)\} + \pi_{1j}\{1 - Bin(n_j\lambda_{1j}; n_j, \phi_1)\} \\ &\quad + \pi_{2j}Bin(n_j\lambda_{2j} - 1; n_j, \phi_2), \end{aligned} \tag{2}$$

where $Bin(b; n, \phi)$ is the cumulative density function of the binomial distribution, with size and probability parameters n and ϕ evaluated at the value b . It can be shown that the decision error rate $\alpha(\lambda_{1j}, \lambda_{2j})$ is minimized when

$$\lambda_{1j} = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right) + n_j^{-1}\log\left(\frac{\pi_{1j}}{\pi_{0j}}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)} \tag{3}$$

$$\lambda_{2j} = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right) + n_j^{-1}\log\left(\frac{\pi_{0j}}{\pi_{2j}}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)}. \tag{4}$$

The derivation of this result (see the Web Appendix) reveals that λ_{1j} and λ_{2j} have the following interpretation.

Theorem 1. The values of λ_{1j} and λ_{2j} that minimize the local decision error rate (2) are the boundaries at which the posterior probabilities of H_1 and H_2 , respectively, become more likely than that of the H_0 , i.e., $\lambda_{1j} = \text{argmax}_{\hat{p}_j}(\text{pr}(H_1|n_j, m_j) > \text{pr}(H_0|n_j, m_j))$ and $\lambda_{2j} = \text{argmin}_{\hat{p}_j}(\text{pr}(H_2|n_j, m_j) > \text{pr}(H_0|n_j, m_j))$.

This intuitive interpretation of λ_{1j} and λ_{2j} provides a natural justification for our dose assignment rule. That is, we should escalate the dose if $\hat{p}_j \leq \lambda_{1j}$ because the observed data

support that H_1 is more likely than H_0 (i.e., the current dose j is below the MTD); and we should deescalate the dose if $\hat{p}_j \geq \lambda_{2j}$ because the observed data support that H_2 is more likely than H_0 (i.e., the current dose j is above the MTD).

In addition, if we assign equal prior probabilities to three hypotheses (i.e., $\pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$), the values of λ_{1j} and λ_{2j} simply become the likelihood-ratio hypothesis-testing boundaries. In this case, the value of λ_{1j} is always located between ϕ_1 and ϕ and the value of λ_{2j} is always located between ϕ and ϕ_2 (i.e., $\phi_1 < \lambda_{1j} < \phi$ and $\phi < \lambda_{2j} < \phi_2$). Figure 1 shows how λ_{1j} and λ_{2j} change with respect to ϕ_1 and ϕ_2 when $\phi = 0.25$. We can see that when ϕ_1 and ϕ_2 are symmetric around ϕ (i.e., $\delta_1 = \delta_2$), the resulting λ_{1j} and λ_{2j} are also close to, although not exactly symmetric around ϕ . Interestingly, based on equations (3) and (4), it is easy to see that when $\pi_{0j} = \pi_{1j} = \pi_{2j}$, λ_{1j} and λ_{2j} are invariant to both dose level j and the accumulative sample size n_j . This property is appealing in practice because in this case the same pair of interval boundaries can be conveniently used throughout the trial, regardless of the dose level and the number of patients treated in the trial, which thus greatly simplifies the trial conduct.

2.2.2. Design properties

Cheung (2005) introduced the concept of coherence for trial design and defined it as the design property by which dose escalation (or deescalation) is prohibited when the observed toxicity rate in the most recently treated cohort is more (or less) than the target toxicity rate. Because that definition is based on the response from only the most recently treated cohort, ignoring the responses from cohorts previously enrolled and treated, herein we refer to it as *short-term memory coherence*. As an extension, we define *long-term memory coherence* as the design property by which dose escalation (or deescalation) is prohibited when the observed toxicity rate in the accumulative cohorts at the current dose is more (or less) than the target toxicity rate. From a practical view point, long-term memory coherence is a more relevant design property for conducting clinical trials than short-term memory coherence. This is because when clinicians determine whether a dose escalation (or deescalation) is practically plausible, they almost always base their decision on the toxicity data from all patients, rather than the most recently treated cohort (i.e., rarely more than 3 patients), treated at that dose. As shown in the Web Appendix, the local BOIN design has the finite-sample property given in Theorem 2.

Theorem 2. When setting $\pi_{0j} = \pi_{1j} = \pi_{2j}$, the local BOIN design is long-term memory coherent in the sense that the probability of dose escalation (or deescalation) is zero when the observed toxicity rate \hat{p}_j at the current dose is higher (or lower) than the target toxicity rate ϕ .

This finite-sample property makes the local BOIN design very appealing in practice because it automatically satisfies the following (ad hoc) safety requirement often imposed by clinicians: dose escalation is not allowed if the observed toxicity rate at the current dose is higher than the MTD. Because the local BOIN design uses the toxicity information from previously treated patients to make the decision of dose transition, it is not short-term memory coherent. We note that the commonly used version of the CRM, which does not allow a dose to be skipped, is not short-term memory coherent either, especially at the beginning of the trial and when using an informative prior.

We now turn to the large-sample property of the design. Under the proposed optimal dose assignment, we tend to treat patients at (or close to) the MTD, which facilitates the MTD selection at the end of the trial because most data and statistical power are concentrated around the MTD. Since $\lim_{n_j \rightarrow \infty} \lambda_{1j} = \log(\frac{1-\phi_1}{1-\phi})/\log(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}) \equiv \tilde{\lambda}_1$ and $\lim_{n_j \rightarrow \infty} \lambda_{2j} = \log(\frac{1-\phi}{1-\phi_2})/\log(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}) \equiv \tilde{\lambda}_2$, it can be shown that the proposed design has the asymptotic dose-selection property described in Theorem 3.

Theorem 3. Dose allocation in the local BOIN design converges almost surely to dose level j^* if $p_{j^*} \in (\tilde{\lambda}_1, \tilde{\lambda}_2)$ and dose level j^* is the only dose satisfying $p_{j^*} \in [\tilde{\lambda}_1, \tilde{\lambda}_2]$. If no dose level satisfies $p_j \in (\tilde{\lambda}_1, \tilde{\lambda}_2)$ but $\phi \in [p_1, p_J]$, the local BOIN design would eventually oscillate almost surely between the two dose levels at which the associated toxicity probabilities straddle the target interval. If there are multiple dose levels satisfying $p_j \in (\tilde{\lambda}_1, \tilde{\lambda}_2)$, the local BOIN design will converge almost surely to one of these levels.

The proof of this result is similar to that of Oron, Azriel and Hoff (2011), noting that λ_{1j} and λ_{2j} converge to constants $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$.

2.2.3. Practical implementation

To use the local BOIN designs in practice, we need to specify the values of ϕ_1 and ϕ_2 . In general, we should avoid setting the values of ϕ_1 and ϕ_2 very close to ϕ . This is because the small sample sizes of typical phase I trials prevent us from differentiating the target toxicity rate from the rates close to it. In addition, in most clinical applications, the target toxicity rate is often a rough guess, and finding a dose level with a toxicity rate reasonably close to the target rate will still prove to be of interest to the investigator. Based on our experience with phase I oncology trials and the operating characteristics of the proposed design, we find that $\phi_1 \in [0.5\phi, 0.7\phi]$ and $\phi_2 \in [1.3\phi, 1.5\phi]$ are reasonable values for most clinical applications. As default values, we recommend $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$ (i.e., 40% deviation from the target) for general use. For example, if the target toxicity is 0.25, we may set $\phi_1 = 0.15$ and $\phi_2 = 0.35$.

The other parameters we need to specify for the trial design are π_{kj} , the prior probability of the hypothesis H_{kj} , $k = 0, 1, 2$. As the π_{kj} 's represent the prior probabilities that dose level j is below, equal to or above the MTD, their values can be directly elicited from physicians by asking them how likely it is that each of J doses is below, equal to or above the MTD. When such prior information is not available, as is often the case in practice, we can take a “noninformative” approach by assuming that the current dose has equal prior probabilities of being below, equal to or above the MTD. Since the interval design examines the dose levels one at a time (i.e., the current dose level) without borrowing information from other dose levels, and the current dose level can be at any level from 1 to J , this noninformative approach effectively results in the noninformative prior $\pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$. One advantage of using such a noninformative prior is that the resulting interval boundaries λ_{1j} and λ_{2j} are invariant to dose level j and the number of patients n_j . Therefore, the same pair of boundaries can be conveniently used throughout the trial for dose assignment, regardless of the dose level and the number of enrolled patients. Table 1 provides the interval boundaries for some commonly encountered target toxicity rates in oncology under the noninformative prior $\pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$. It seems appealing to specify increasing prior probabilities for π_{2j} (i.e., $\pi_{21} < \pi_{22} < \dots < \pi_{2J}$) to reflect that

toxicity monotonically increases with dose levels. However, based on numerical studies, we found that such an approach did not improve the performance of the design (results are not shown). This may be because the monotonicity of the dose-toxicity relationship has been (implicitly) incorporated into the design through the dose escalation rule (i.e., we escalate the dose because we assume that the next higher dose level is more toxic), and thus using ordered priors will not bring enough extra information to improve the design performance.

The local BOIN design is very easy to implement in practice. Once we specify the design parameters as described above, the interval boundaries λ_{1j} and λ_{2j} can be easily calculated at the trial design phase based on equations (3) and (4). Then during the trial conduct, clinicians can simply count the number of patients who experience toxicity and compare the observed toxicity rate \hat{p}_j with the prespecified interval boundaries λ_{1j} and λ_{2j} to determine dose assignment until the trial is completed.

After the trial is completed, we need to select a dose level as the MTD. We propose to select the MTD based on $\{\tilde{p}_j\}$, the isotonically transformed values of the observed toxicity rates $\{\hat{p}_j\}$. Specifically, we select as the MTD dose j^* , for which the isotonic estimate of toxicity rate \tilde{p}_{j^*} is closest to ϕ ; if there are ties for \tilde{p}_{j^*} , we select from the ties the highest dose level when $\tilde{p}_{j^*} < \phi$ or the lowest dose level when $\tilde{p}_{j^*} > \phi$. The isotonic estimates $\{\tilde{p}_j\}$ can be obtained by applying the pooled-adjacent-violators algorithm (PAVA) (Barlow et al., 1972) to $\{\hat{p}_j\}$. Operatively, the PAVA replaces any adjacent \hat{p}_j 's that violate the nondecreasing order by their (weighted) average so that the resulting estimates \tilde{p}_j 's become monotonic. In the case in which the observed toxicity rates are monotonic, \tilde{p}_j and \hat{p}_j are equivalent. We note that besides the above method, any other reasonable dose selection procedure can be used to select the MTD for the local optimal design, as the MTD selection and dose transition rules are two relatively independent components of the trial design, both conceptually and operatively.

One practical issue largely ignored in the previously proposed interval designs involves the risk of assigning too many patients to an overly toxic dose. Because interval designs do not “look” ahead, but instead use only the toxicity information at the current dose level to determine the dose escalation decision, the dose assignment will bounce back and forth between two adjacent doses when one of them is much lower than the MTD and the other is much higher than the MTD. To avoid this result of assigning too many patients to the overly toxic dose, we propose to impose the following dose elimination rule when implementing the local BOIN design.

If $\text{pr}(p_j > \phi | m_j, n_j) > 0.95$ and $n_j \geq 3$, dose levels j and higher are eliminated
from the trial, and the trial is terminated if the first dose level is eliminated,

where $\text{pr}(p_j > \phi | m_j, n_j) > 0.95$ can be evaluated based on a beta-binomial model, assuming that m_j follows a binomial distribution (with size and probability parameters n_j and p_j) and p_j follows a vague beta prior, e.g., $p_j \sim \text{beta}(1, 1)$. Based on our experience, rather than repeatedly evaluating the above dose elimination rule in real time during the trial conduct, medical researchers often prefer to enumerate the dose elimination boundaries for each possible value of n_j before the initiation of the trial and include these boundaries in the trial protocol. Therefore, when conducting the trial, they can determine dose elimination by simply examining whether the number of patients experiencing toxicity at the current dose, i.e., m_j , exceeds the elimination boundaries. Table 2 (in the bottom row) provides the elimination boundaries for $\phi = 0.25$. For example, when the number of patients treated at the current dose $n_j = 4$, we need to eliminate that dose and higher doses if 3 or more

patients experience toxicity.

To facilitate practitioners applying the proposed designs, we have prepared easy-to-use software (written in R with detailed instructions) to calculate the interval and dose elimination boundaries and select the MTD. The software is available for free download at <http://odin.mdacc.tmc.edu/~yyuan/>.

2.3. Global BOIN design

We have discussed the local BOIN design, which minimizes the decision error under the three point hypotheses given by equation (1). In this section, we discuss a global BOIN design, which accounts for all possible values of p_j by specifying three composite hypotheses. In the global BOIN design, values of λ_{1j} and λ_{2j} are chosen to minimize the average decision error over the whole support of p_j , i.e., $p_j \in [0, 1]$. The contrast between the local and global optimal interval designs is somewhat analogous to the uniformly most powerful tests for simple hypotheses versus composite hypotheses in the frequentist testing framework. Specifically, we define three composite hypotheses,

$$\begin{aligned} H_{0j} &: \phi_1 < p_j < \phi_2 \\ H_{1j} &: 0 \leq p_j \leq \phi_1 \\ H_{2j} &: \phi_2 \leq p_j \leq 1, \end{aligned}$$

where H_{1j} indicates that dose level j is subtherapeutic and we should escalate the dose; H_{2j} indicates that dose level j is overly toxic and we should deescalate the dose; and H_{0j} indicates that dose level j is close to the MTD and we should retain the same dose level.

Under each hypothesis, we assign the toxicity probabilities p_j a noninformative uniform prior: $f(p_j|H_{0j}) = \text{Unif}(\phi_1, \phi_2)$, $f(p_j|H_{1j}) = \text{Unif}(0, \phi_1)$, and $f(p_j|H_{2j}) = \text{Unif}(\phi_2, 1)$. Then the global decision error rate is given by

$$\begin{aligned} \alpha_g(\lambda_{1j}, \lambda_{2j}) &= \text{pr}(H_{0j}) \int f(p_j|H_{0j}) \text{pr}(\bar{\mathcal{R}}|p_j, H_{0j}) dp_j + \text{pr}(H_{1j}) \int f(p_j|H_{1j}) \text{pr}(\bar{\mathcal{E}}|p_j, H_{1j}) dp_j \\ &\quad + \text{pr}(H_{2j}) \int f(p_j|H_{2j}) \text{pr}(\bar{\mathcal{D}}|p_j, H_{2j}) dp_j \tag{5} \\ &= \pi_{0j} + \pi_{1j} + \sum_{y=0}^{b_{1j}} \left[\frac{\pi_{0j} \{ \text{Beta}(\phi_2; y+1, n_j - y + 1) - \text{Beta}(\phi_1; y+1, n_j - y + 1) \}}{(\phi_2 - \phi_1)(n_j + 1)} \right. \\ &\quad \left. - \frac{\pi_{1j} \text{Beta}(\phi_1; y+1, n_j - y + 1)}{\phi_1(n_j + 1)} \right] + \sum_{y=0}^{b_{2j}-1} \left\{ \frac{\pi_{2j} \text{Beta}(\phi_2; y+1, n_j - y + 1)}{(1 - \phi_2)(n_j + 1)} \right. \\ &\quad \left. - \frac{\pi_{0j} \{ \text{Beta}(\phi_2; y+1, n_j - y + 1) - \text{Beta}(\phi_1; y+1, n_j - y + 1) \}}{(\phi_2 - \phi_1)(n_j + 1)} \right\}, \end{aligned}$$

where $b_{1j} = \text{floor}(n_j \lambda_{1j})$, $b_{2j} = \text{floor}(n_j \lambda_{2j})$ and $\text{Beta}(c; a, b)$ is the cumulative distribution function of a beta distribution with the shape and scale parameters a and b , evaluated at the value c . Although without closed forms, the following theorem can be shown to hold.

Theorem 4. The values of λ_{1j} and λ_{2j} that minimize the global decision error rate (5) are the boundaries at which the posterior probabilities of H_1 and H_2 , respectively, become

more likely than that of H_0 .

The proof of this result appears in the Web Appendix. In practice, the values of λ_{1j} and λ_{2j} that minimize the global decision error rate can be easily determined by a numerical search, as b_{1j} and b_{2j} are integers between 0 and n_j .

Compared to the local BOIN design, the interval length (i.e., $\lambda_{2j} - \lambda_{1j}$) under the global BOIN design is wider (see Figure 2). In addition, as shown in Figure 2, unlike the local BOIN design, for which λ_{1j} and λ_{2j} are invariant to n_j , for the global BOIN design, λ_{1j} and λ_{2j} depend on n_j even when the three hypotheses have equal prior probabilities. Despite these differences, we note that the global BOIN design is also long-term memory coherent because $\lambda_{1j} < \phi$ and $\lambda_{2j} > \phi$.

Theorem 5. When setting $\pi_{0j} = \pi_{1j} = \pi_{2j}$, the global BOIN design is long-term memory coherent in the sense that the probability of dose escalation (or deescalation) is zero when the observed toxicity rate \hat{p}_j at the current dose is higher (or lower) than the target toxicity rate ϕ .

In addition, as shown in the Web Appendix, the global BOIN design has the convergence property described in Theorem 6.

Theorem 6. Dose allocation in the global BOIN design converges almost surely to dose level j^* if $p_{j^*} \in (\phi_1, \phi_2)$ and dose level j^* is the only dose satisfying $p_{j^*} \in [\phi_1, \phi_2]$. If no dose level satisfies $p_j \in (\phi_1, \phi_2)$ but $\phi \in [p_1, p_J]$, the global BOIN design would eventually oscillate almost surely between the two dose levels at which the associated toxicity probabilities straddle the target interval. If there are multiple dose levels satisfying $p_j \in (\phi_1, \phi_2)$, the global BOIN design will converge almost surely to one of these levels.

As for the practical implementation of this design, the same principle described in the previous section can be used to specify the design parameters ϕ_1 , ϕ_2 and π_{kj} for the global BOIN design. The interval boundaries for the global BOIN design do not have a closed form (but can be easily determined using a numerical search) and depend on n_j even when we set $\pi_{0j} = \pi_{1j} = \pi_{2j}$. This makes the implementation of the global BOIN design slightly more complicated than that of the local optimal design. That is, we need to find λ_{1j} and λ_{2j} for each possible value of n_j before the initiation of the trial. Once these boundaries are determined, the global BOIN design can be conducted by comparing the observed toxicity rate with the interval boundaries. Table 2 displays the interval boundaries of the global BOIN design when the target toxicity rate is 0.25 up to $n_j = 15$. Thus for practical reasons, we may prefer to use the local BOIN design. The finite sample simulation described in the next section also suggests that the local BOIN design has better operating characteristics.

3. Simulation

We used simulation studies to compare the operating characteristics of the BOIN designs with those of three available designs: the cumulative cohort design (CCD), the modified toxicity probability interval (mTPI) design, and the CRM. We assumed a total of $J = 6$ dose levels with a maximum sample size of $n = 36$ patients in 12 cohorts of size 3. The target toxicity rate was $\phi = 0.25$. In the proposed designs, we set $\phi_1 = 0.15$, $\phi_2 = 0.35$

and assigned equal prior probabilities to three hypotheses with $\pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$. The same values of ϕ_1 and ϕ_2 were also used for the mTPI design. For the CCD, following Ivanova et al. (2007), we set the tolerance interval as $\phi \pm 0.09$. The CRM was based on the model $p_j = a_j^{\exp(\alpha)}$ where the “skeleton” $a_j = (0.01, 0.08, 0.25, 0.46, 0.65, 0.79)$ was chosen based on the model calibration method of Lee and Cheung (2009). We assigned the unknown parameter α the least-informative prior $\alpha \sim N(0, 1.24^2)$ as proposed by Lee and Cheung (2011), under which each dose level has an (approximately) equal prior probability of being the MTD. Skipping a dose level was not allowed in the CRM. To make the designs more comparable, we applied the dose elimination rule and isotonic dose selection rule, as described in Section 2.2.3, to all designs except the CRM because that design has its own model-based safety and dose selection rules. We simulated 10,000 trials. To avoid cherry-picking scenarios that are better suited to specific methods, for each simulated trial, the toxicity scenario (i.e., the true toxicity probabilities of the six doses) was randomly generated based on the approach of Paoletti, O’Quigley and Maccario (2004) as follows:

- (a) Randomly select, with equal probabilities, one of the J dose levels as the MTD and denote this selected dose level by j .
- (b) Generate the toxicity probability of the MTD, $p_j = \Phi(\epsilon_j)$, where the random error $\epsilon_j \sim N(z(\phi), \sigma_0^2)$, with $\Phi(\cdot)$ and $z(\cdot)$ denoting the cumulative density function (CDF) and the inverse CDF of the standard normal distribution, respectively. Here we allow the true toxicity probability of the MTD to deviate from ϕ , as is often the case in practice.
- (c) Generate p_{j-1} and p_{j+1} (i.e., the toxicity probabilities of two doses adjacent to the MTD) under the constraint that p_j is closest to ϕ . This can be done by generating $p_{j-1} = \Phi[z(p_j) - \{z(p_j) - z(2\phi - p_j)\}I(z(p_j) > z(\phi)) - \epsilon_{j-1}^2]$ and $p_{j+1} = \Phi[z(p_j) + \{z(2\phi - p_j) - z(p_j)\}I(z(p_j) < z(\phi)) + \epsilon_{j+1}^2]$ where $\epsilon_{j-1} \sim N(\mu_1, \sigma_1^2)$, $\epsilon_{j+1} \sim N(\mu_2, \sigma_2^2)$, and $I(\cdot)$ is an indicator function with $I(z(p_j) > z(\phi)) = 1$ if $z(p_j) > z(\phi)$ and 0 otherwise.
- (d) Successively generate the toxicity probabilities for the remaining levels according to $p_{j-2} = \Phi[z(p_{j-1}) - \epsilon_{j-2}^2]$ and $p_{j+2} = \Phi[z(p_{j+1}) + \epsilon_{j+2}^2]$, and $p_{j-3} = \Phi[z(p_{j-2}) - \epsilon_{j-3}^2]$ and $p_{j+3} = \Phi[z(p_{j+2}) + \epsilon_{j+3}^2]$, and so on, where $\epsilon_{j-2}, \epsilon_{j-3} \sim N(\mu_1, \sigma_1^2)$ and $\epsilon_{j+2}, \epsilon_{j+3} \sim N(\mu_2, \sigma_2^2)$.

In our simulation, we set $\sigma_0 = 0.05$, $\sigma_1 = \sigma_2 = 0.35$, and we controlled the average probability difference around the target to be 0.07, 0.1 and 0.15 by varying the values of μ_1 and μ_2 . For illustration, Figure 3 shows 10 (out of 10,000) randomly generated dose-toxicity curves when the average probability difference around the target was 0.1. We can see that these curves demonstrate various shapes and locations of the MTD.

Table 3 shows the simulation results, including the selection percentage of the MTD, the average percentage of patients treated at the MTD, the average toxicity rate, and the average sample size. In addition, we reported two risk measures for the designs: (1) *the risk of poor allocation*, defined as the percentage of simulation runs in which the number of patients allocated to the MTD (say n_{MTD}) is less than that of a standard non-sequential design, which assigns the equal number of patients to each dose, i.e., $\Pr(n_{\text{MTD}} < n/J)$; and (2) *the risk of high toxicity*, defined as the percentage of simulation runs in which the total number of toxicities is greater than $n\phi$. These risk measures are of great practical importance because they gauge the likelihood of a trial turning out to be a “bad” trial (e.g., performing worse than a standard non-sequential trial) under a specific design. In other

words, they measure the reliability (or variation) of the design. This important aspect of trial design has been largely overlooked by the existing literature, which typically focuses only on the mean or average performance of a design.

The simulation results show that all the designs generally have a similar “average” level of performance regarding several factors, including the MTD selection percentage, the average percentage of patients treated at the MTD, and the average toxicity rate. However, in terms of the risk of poor allocation decisions, the local BOIN design stands out and outperforms the other competing designs. For example, under the local BOIN design, the risk of poor allocation is about 14% to 16% lower than that under the CRM, and is about 11% lower than those under the CCD and mTPI when the average probability difference around the target is 0.1 and 0.07. For the risk of high toxicity, the proposed local BOIN design also performs better than the other designs. These results suggest that the local BOIN design has better reliability, and the use of the local BOIN design decreases the likelihood of conducting a poorly performing trial compared to the use of the other designs. Interestingly, compared to the local BOIN design, the proposed global BOIN design had a higher risk of poor allocation and high toxicity. This is because by its construction, the global BOIN design “wastes” its power to accommodate the cases that are (virtually) practically irrelevant, for example, to minimize the decision error between $\phi_1 + \nu$ (i.e., H_0 is true) and $\phi_1 - \nu$ (i.e., H_1 is true), and between $\phi_2 + \nu$ (i.e., H_2 is true) and $\phi_2 - \nu$ (i.e., H_0 is true), where ν is a very small number.

4. Application

We applied the BOIN designs to the aforementioned solid tumor trial. We elicited from physicians four representative dose-toxicity scenarios that we may encounter in practice. These scenarios had different locations for the MTD. Under each scenario, we simulated 10,000 trials to evaluate the operating characteristics of the BOIN design for the purpose of trial protocol preparation.

Table 4 shows the operating characteristics of the BOIN designs. For comparison, we also report the results for the CCD, mTPI and CRM designs. The design parameters of these designs were the same as described previously. Compared to these designs, the local BOIN design yielded roughly comparable “average” performance levels (i.e., the MTD selection percentage and the average number of patients allocated to the MTD), but generally had substantially lower risks of poor allocation decisions or high toxicity. For example, in scenario 1, the first dose was the MTD, and the local BOIN design yielded the highest selection percentage of 63%. The risk of poor allocation for the local BOIN design was only about 1/3 or 1/4 of those of the other designs. Scenario 2 had the second dose as the MTD, and the local BOIN design selected the target dose 66.3% of the time, which was similar to the target selection percentages of the CCD and mTPI designs. However, the risk of poor allocation under the local BOIN design was less than half of those under the CCD and mTPI designs, and the risk of high toxicity under the local BOIN design was also about 6-9% lower. We note that the relative performance of the CRM depended on how much the assumed model structure (e.g., skeleton) deviated from the truth. The CRM performed better than the other designs if the target dose was close to its prior estimate of the MTD (e.g., under scenarios 2 and 3), but performed worse than the other designs if the target dose deviated from its prior estimate (e.g., under scenarios 1 and 4).

As a second application, we retrospectively applied the local BOIN design to the pedi-

atric phase I clinical trial (Jakacki et al., 2008). Given the target toxicity rate $\phi = 0.25$, the dose escalation and deescalation boundaries were $\lambda_{1j} = 0.197$ and $\lambda_{2j} = 0.298$ (see Table 1). The trial started with treating the first cohort of three patients at the lowest dose, 35 mg, and no DLT was observed. As the observed toxicity rate was $\hat{p}_1 = 0$, lower than the escalation boundary λ_{1j} , we escalated the dose to 50 mg for the second cohort. Again, no DLT was observed with $\hat{p}_2 = 0$, so the dose was escalated to 65 mg, at which still no DLT occurred. The dose was continuously escalated to 85 mg, at which one out of six patients experienced DLT. At that moment, \hat{p}_4 was equal to $1/6$, but was still smaller than the escalation boundary $\lambda_{1j} = 0.197$, and thus the dose was escalated to 110 mg. The trial ended by treating the last cohort of patients at a dose of 110 mg, at which two out of four patients experienced DLTs. Based on the observed data and applying the PAVA algorithm, the isotonic estimates of the toxicity probabilities of the five doses were 0.002, 0.002, 0.002, 0.167 and 0.500, respectively. As a result, dose level 4 was selected as the MTD.

5. Conclusion

We have proposed a flexible framework for constructing interval designs by treating phase I dose finding as a decision-making problem. We proposed local and global BOIN designs that minimize the decision error of dose assignment based on either point or composite hypotheses. We showed that the BOIN designs have sound theoretical properties and good numerical performance. Among two proposed designs, we recommend the local BOIN design for practical use because of its simplicity and better finite-sample performance.

We have considered the local and global optimal designs, but the proposed decision-making framework is very flexible. These designs can be easily modified to accommodate different design objectives by specifying an appropriate object function to be minimized. For example, noting that the local and global decision errors given in equations (2) and (5) actually consist of different types of decision errors, we can propose minimax designs to prevent the rate of any individual type of decision error from being too high.

In some cases, for safety reasons, we may be more concerned with incorrect dose escalation than with incorrect dose deescalation. In these cases, we can classify the decision error, such as equation (2), into errors of making incorrect decisions of escalation, deescalation and dose level retainment. We then assign the appropriate weight to each type of error to reflect its relative importance, and minimize the resulting objective function.

ACKNOWLEDGEMENTS

The authors would like to thank the editor, associate editor and reviewers for insightful and constructive comments. Yuan's research was supported by Award Number R01 CA154591 and P50 CA098258 from the National Cancer Institute.

References

- Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972) *Statistical Inference under Order Restrictions*, London, New York: John Wiley & Sons.
- Cheung, Y. (2005) Coherence principles in dose-finding studies. *Biometrika*, **92**, 863-873.
- Chevret, S. (2006) *Statistical Methods for Dose-Finding Experiments*, London: John Wiley & Sons.
- Durham, S.D. and Flournoy, N. (1994) Random walks for quantile estimation. In *Statistical Decision Theory and Related Topics* (eds., J.O. Berger, S. S. Gupta), pp. 467-476. New York: Springer.
- Gerke, O. and Siedentop, H. (2008) Optimal phase I dose-escalation trial designs in oncology—a simulation study. *Stat. Med.*, **27**, 5329-5344.
- Ivanova, A., Flournoy, N., and Chung, Y. (2007) Cumulative cohort design for dosefinding. *J. Stat. Plan. Inference*, **137**, 2316-2317.
- Ivanova, A. and Kim, S. (2009) Dose-finding for binary ordinal and continuous outcomes with monotone objective function: a unified approach. *Biometrics*, **65**, 307-315.
- Jakacki, R. I., Hamilton, M., Gilbertson, J. R., Blaney, S. M., Tersak, J., Krailo, M. D., Ingle, A. M., Voss, S. D., Dancey, J. E., and Adamson, P. C. (2008), Pediatric phase I and pharmacokinetic study of erlotinib followed by the combination of erlotinib and temozolomide: a Childrens Oncology Group Phase I Consortium Study, *J. Clin. Oncol.*, **26**, 4921- 4927.
- Ji, Y., Liu, P., Li, Y., and Bekele, B. N. (2010) A modified toxicity probability interval method for dose-finding trials. *Clin. Trials*, **7**, 653-663.
- Korn, E.L., Midthune, D., Chen, T.T., Rubinstein, L.V., Christian, M.C., and Simon, R.M. (1994) A comparison of two phase I trial designs. *Stat. Med.*, **13**, 1799-1806.
- Lee, S. and Cheung, Y. (2009) Model calibration in the continual reassessment method. *Clin. Trials*, **6**, 227-238.
- Lee, S. and Cheung, Y. (2011) Calibration of prior variance in the Bayesian continual reassessment method. *Stat. Med.*, **30**, 2081-2089.
- Leung, D. and Wang, Y.G. (2001) Isotonic designs for phase I trials. *Control. Clin. Trials*, **22**, 126-138.
- Lin, Y. and Shih, W.J. (2001) Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. *Biostatistics*, **2**, 203-215.
- Neuenschwander B., Branson M. and Gsponer T. (2008) Critical aspects of the Bayesian approach to phase I cancer trials. *Stat. Med.*, **27**, 2420-2439.
- Oron, A., Azriel, D. and Hoff, P. (2011) Dose-finding designs: the role of convergence properties. *Int. J. Biostat.*, **7**, Article 39.

- O'Quigley, J., Pepe, M., and Fisher, L. (1990) Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*, **46**, 33-48.
- Paoletti, X. , O'Quigley, J. and Maccario, J. (2004) Design efficiency in dose finding studies. *Comput. Stat. Data Anal.*, **45**, 197-214.
- Storer, B. E. (1989) Design and analysis of phase I clinical trials. *Biometrics*, **45**, 925-937.
- Ting, N. (2006) *Dose Finding in Drug Development*, Cambridge: Springer.
- Walker, S.G. and Hjort, N.L. (2001) On Bayesian consistency. *J. R. Stat. Soc. Series B Stat. Methodol.*, **63**, 811-821.
- Whitehead, J. and Brunier, H. (1995) Bayesian decision procedures for dose determining experiments. *Stat. Med.*, **14**, 885-893.
- Yuan, Z. and Chappell, R. (2004) Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clin. Trials*, **1**, 499-508.

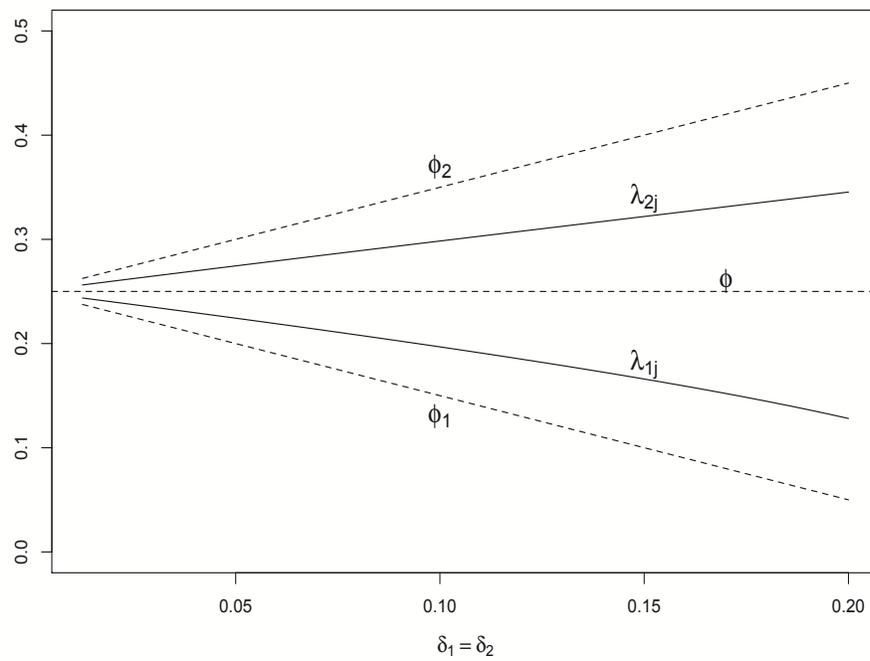


Fig. 1. Interval boundaries λ_{1j} and λ_{2j} for the local BOIN design change with the values of ϕ_1 and ϕ_2 when the target toxicity rate $\phi = 0.25$ and $\pi_{0j} = \pi_{1j} = \pi_{2j}$.

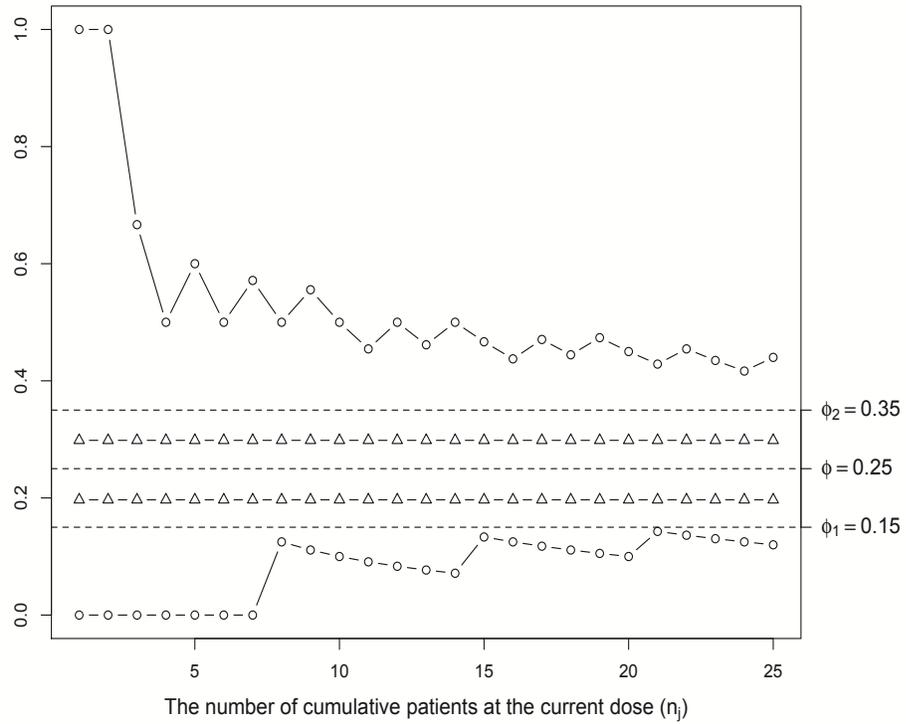


Fig. 2. Interval boundaries λ_{1j} and λ_{2j} for the local BOIN design (triangles) and global BOIN design (circles) under different numbers of cumulative patients at the current dose (n_j) when $\phi = 0.25$, $\phi_1 = 0.15$, $\phi_2 = 0.35$ and $\pi_{0j} = \pi_{1j} = \pi_{2j}$. Under each design, the top line is λ_{1j} and the bottom line is λ_{2j} .

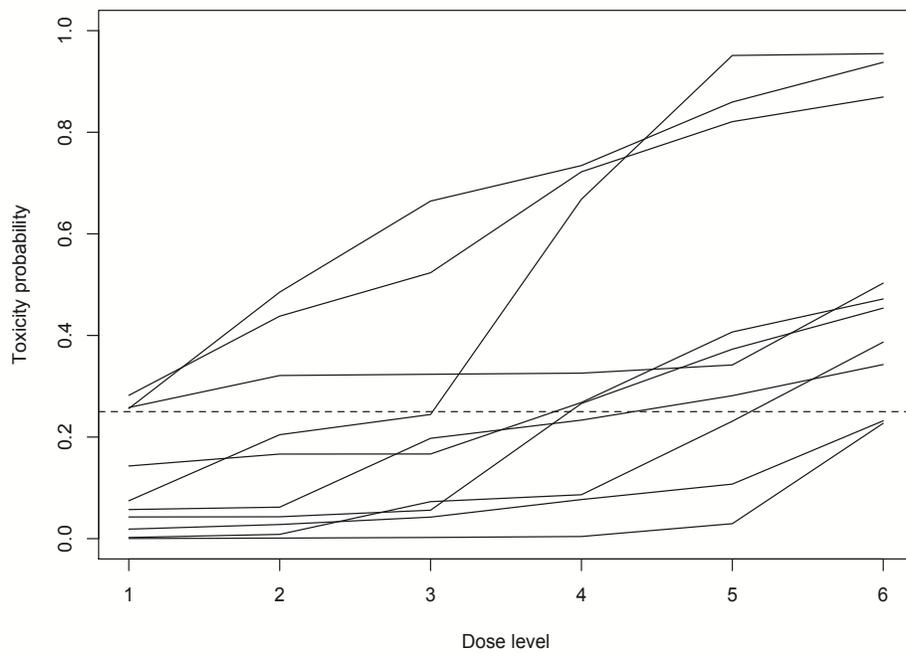


Fig. 3. Ten randomly generated dose-toxicity curves based on the simulation procedure described in Section 3. The dashed line indicates the target toxicity probability of 0.25.

Table 1. Values of λ_{1j} and λ_{2j} under the local BOIN design for different target toxicity rates with $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$.

Interval boundaries	Target toxicity rate ϕ					
	0.15	0.2	0.25	0.3	0.35	0.4
λ_{1j}	0.118	0.157	0.197	0.236	0.276	0.316
λ_{2j}	0.179	0.238	0.298	0.358	0.419	0.479

Table 2. Interval boundaries and dose elimination boundaries for the global BOIN design when the target toxicity rate $\phi = 0.25$ with $\phi_1 = 0.15$ and $\phi_2 = 0.35$.

Boundary	Number of cumulative patients treated at the current dose (n_j)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
λ_{1j}	0/1	0/2	0/3	0/4	0/5	0/6	0/7	1/8	1/9	1/10	1/11	1/12	1/13	1/14	2/15
λ_{2j}	1/1	2/2	2/3	2/4	3/5	3/6	4/7	4/8	5/9	5/10	5/11	6/12	6/13	7/14	7/15
Elimination	NA	NA	3/3	3/4	3/5	4/6	4/7	4/8	5/9	5/10	6/11	6/12	6/13	7/14	7/15

Table 3. Simulation results when the dose-toxicity scenarios are randomly generated.

Design	Selection % of MTD	% of patients at MTD	Average toxicity rate	Risk of poor allocation	Risk of high toxicity	Sample size
Average probability difference around the target = 0.1						
CCD	46.0	35.2	19.8	39.1	16.2	35.3
mTPI	44.8	35.3	19.8	40.2	17.3	35.2
CRM	44.7	35.4	19.3	43.0	16.9	35.2
Global BOIN	45.4	35.5	19.9	41.7	17.6	35.2
Local BOIN	46.2	33.0	19.2	28.7	14.8	35.3
Average probability difference around the target = 0.07						
CCD	36.2	29.6	20.0	49.4	16.5	35.1
mTPI	36.1	29.8	20.0	50.7	17.6	35.1
CRM	33.3	28.6	19.5	54.7	17.5	35.1
Global BOIN	36.6	29.6	20.1	51.5	18.0	35.1
Local BOIN	37.0	28.5	19.5	38.6	15.5	35.1
Average probability difference around the target = 0.15						
CCD	57.7	43.4	19.8	26.0	16.8	35.4
mTPI	57.6	44.0	20.1	27.1	18.1	35.4
CRM	59.8	44.7	19.6	26.6	17.9	35.4
Global BOIN	57.5	44.1	20.1	27.7	18.1	35.4
Local BOIN	57.6	43.4	19.0	18.0	15.5	35.3

Table 4. Operating characteristics of the BOIN designs under four prespecified dose-toxicity scenarios for the solid tumor trial. The target toxicity rate of 0.25 is underlined.

Design		Dose level						Risk of poor allocation	Risk of high toxicity
		1	2	3	4	5	6		
Scenario 1	Pr(toxicity)	<u>0.25</u>	0.35	0.5	0.6	0.7	0.8		
CCD	Selection (%)	60.8	23.0	1.5	0.0	0.0	0.0	14.7	52.8
	# patients	22.4	8.8	1.4	0.1	0.0	0.0		
mTPI	Selection (%)	58.2	25.3	1.8	0.1	0.0	0.0	19.6	53.7
	# patients	21.8	9.3	1.5	0.1	0.0	0.0		
CRM	Selection (%)	55.6	26.2	1.1	0.0	0.0	0.0	15.9	52.9
	# patients	21.9	9.3	1.4	0.1	0.0	0.0		
Global BOIN	Selection (%)	59.4	24.5	1.6	0.0	0.0	0.0	19.7	54.1
	# patients	21.5	9.5	1.5	0.1	0.0	0.0		
Local BOIN	Selection (%)	63.0	20.6	1.6	0.1	0.0	0.0	5.9	53.4
	# patients	22.6	8.3	1.7	0.2	0.0	0.0		
Scenario 2	Pr(toxicity)	0.1	<u>0.25</u>	0.4	0.6	0.7	0.8		
CCD	Selection (%)	17.7	65.7	15.7	0.6	0.0	0.0	14.1	31.8
	# patients	10.4	18.4	6.3	0.7	0.0	0.0		
mTPI	Selection (%)	16.9	64.0	18.1	0.7	0.0	0.0	18.3	34.8
	# patients	9.9	18.4	6.7	0.7	0.0	0.0		
CRM	Selection (%)	10.9	74.1	14.6	0.1	0.0	0.0	12.0	35.2
	# patients	9.6	19.5	6.4	0.4	0.0	0.0		
Global BOIN	Selection (%)	18.0	63.8	17.3	0.6	0.0	0.0	19.1	36.5
	# patients	9.7	18.5	6.9	0.8	0.0	0.0		
Local BOIN	Selection (%)	18.4	66.3	14.4	0.5	0.0	0.0	5.7	25.8
	# patients	12.4	16.4	6.1	0.9	0.1	0.0		
Scenario 3	Pr(toxicity)	0.05	0.1	<u>0.25</u>	0.32	0.5	0.6		
CCD	Selection (%)	0.5	18.8	50.4	27.4	2.8	0.2	26.9	13.2
	# patients	4.5	9.5	14.0	6.5	1.4	0.1		
mTPI	Selection (%)	0.5	18.0	49.1	29.0	3.1	0.2	30.5	15.6
	# patients	4.5	9.1	14.1	6.7	1.4	0.1		
CRM	Selection (%)	0.1	18.1	61.1	19.5	1.1	0.0	25.8	14.2
	# patients	4.4	9.7	15.7	5.2	0.8	0.1		
Global BOIN	Selection (%)	0.7	18.3	49.0	28.8	3.0	0.2	31.7	16.8
	# patients	4.5	9.0	14.1	6.9	1.5	0.1		
Local BOIN	Selection (%)	0.4	19.0	53.0	24.7	2.8	0.1	14.5	9.8
	# patients	5.1	10.9	12.4	6.0	1.6	0.2		
Scenario 4	Pr(toxicity)	0.01	0.02	0.03	0.04	0.05	<u>0.25</u>		
CCD	Selection (%)	0.0	0.0	0.1	0.5	15.2	84.2	7.1	0.0
	# patients	3.2	3.4	3.7	3.9	6.2	15.6		
mTPI	Selection (%)	0.0	0.0	0.1	0.5	14.3	85.0	7.3	0.0
	# patients	3.2	3.4	3.7	3.9	6.0	15.8		
CRM	Selection (%)	0.0	0.1	2.1	9.8	18.1	69.9	27.2	0.0
	# patients	3.2	3.5	4.7	5.1	7.0	12.5		
Global BOIN	Selection (%)	0.0	0.0	0.2	0.5	14.5	84.8	7.1	0.0
	# patients	3.2	3.4	3.7	3.9	5.8	16.0		
Local BOIN	Selection (%)	0.0	0.0	0.1	0.7	16.8	82.4	7.4	0.0
	# patients	3.3	3.5	3.8	4.0	8.0	13.4		