An adaptive gBOIN design with shrinkage boundaries for phase I dose-finding trials

Rongji Mu¹, Zongliang Hu²*, Guoying Xu³ and Haitao Pan⁴*

Abstract

Background: With the emergence of molecularly targeted agents and immunotherapies, the landscape of phase I trials in oncology has been changed. Though these new therapeutic agents are very likely induce multiple low- or moderate-grade toxicities instead of DLT, most of the existing phase I trial designs account for the binary toxicity outcomes. Motivated by a pediatric phase I trial of solid tumor with a continuous outcome, we propose an adaptive generalized Bayesian optimal interval design with shrinkage boundaries, gBOINS, which can account for continuous, toxicity grades endpoints and regard the conventional binary endpoint as a special case.

Result: The proposed gBOINS design enjoys convergence properties, e.g., the induced interval shrinks to the toxicity target and the recommended dose converges to the true maximum tolerated dose with increased sample size.

Conclusion: The proposed gBOINS design is transparent and simple to implement. We show that the gBOINS design has the desirable finite property of coherence and large-sample property of consistency. Numerical studies show that the proposed gBOINS design yields good performance and is comparable with or superior to the competing design.

Keywords: Bayesian adaptive design, Phase I dose-finding trial, Shrinkage boundaries, Maximum tolerated dose

Introduction

In oncology phase I trial studies, one main objective is to determine the maximum tolerated dose (MTD) or the recommended phase II dose (RP2D). Targeting on phase I trial studies, numerous methods have been proposed and can be generally classified into three classes: algorithm-based design like the 3+3 design [1], the accelerated titration design [2], and the biased coin design [3]; model-based design like the continual reassessment method (CRM) [4, 5] and its various extensions [6–8]; and recently developed model-assisted design like the Bayesian optimal interval (BOIN) design [9], and the Keyboard design [10]. Note that, all these methods accounting for the binary toxicity outcomes, experienced dose-limiting toxicity (DLT) or not, and thus may not suitable for the trials with multiple low- or moderate-grade toxicities, such as the molecularly targeted or immunotherapy trials [11–14]. To incorporate the toxicity grades (please refer National Cancer Institute Common Terminology Criteria) into dose-finding trials, one way is assigning severity weights to each grade and type of toxicity and combine the weights as a composite score, eg., total toxicity burden (TTB) [15], toxicity burden score (TBS) [16], and total toxicity profile (TTP) [17]. After appropriate transformations, these scores can be taken as the normally distributed variable [14]. The other way is translating toxicity grades to numeric scores which represent their relative severity in the unit of DLT, the ‘equivalent toxicity score’ (ETS) [18] and treating them as quasi-binary end points which take the values ranging from 0 to 1 and can be modelled by the quasi-Bernoulli likelihood [19].

To the best of our knowledge, very few methods in phase I were developed to account for different toxicity
scales, e.g., binary, continuous, count, in a unified framework except a design by [20] and a gBOIN design by [14].

The design by [20] is an algorithm-based design and the gBOIN is a model-assisted design, which is a generalized version of the BOIN design by [9] to account for various toxicity grades. This paper will take further steps to extend the gBOIN. The gBOIN design assumes two fixed boundaries \( \phi_1 \) and \( \phi_2 \), by which the dose transition was conducted. Though gBOIN with fixed boundaries enjoys the desirable performance in finite sample size in the previous study by [14], it leaves the potential for us to improve the performance of gBOIN by exploring its behaviors with non-fixed boundaries. The rationale of studying the non-fixed boundaries is straightforward, since different boundaries are associated with different risks of making wrong dose allocations. In this article, we propose a gBOINS design method, which generalized the gBOIN method with two shrinkage boundaries \( \phi^*_1 \) and \( \phi^*_2 \). These two boundaries are obtained based on the theory of the uniformly most powerful Bayesian test [21]. The trial will be guided by replacing the two fixed boundaries \( \phi_1 \) and \( \phi_2 \) in the gBOIN with \( \phi^*_1 \) and \( \phi^*_2 \), respectively. We show that, in contrast to the gBOIN design which will oscillate among the doses within the equivalent interval, the new proposed gBOINS design has the ideal large-sample behavior that converges to one of the dose levels within the equivalent interval, because its decision boundaries shrink to a point mass toward the target. This distinctive feature of gBOINS provides a theoretical foundation and guarantees the MTD convergence. Numerical studies also show that: for small sample size, the gBOINS yields good performance that is comparable or superior to its raw version gBOIN; for large sample size, compared to the gBOIN design, the performance of gBOINS has a substantial improvement.

Remainder of the paper is organized as follows. In “Method” section, after a brief introduction of gBOIN design we introduce the gBOINS design, its theoretical foundation and derive its properties. In “Simulation” section, we compare the gBOINS design to the gBOIN design with various types of toxicity grades. In “Conclusion” section, we conclude the paper with a discussion.

**Method**

**Introduction of gBOIN design**

Assume there are \( f \) specified doses \( d_1 < \cdots < d_f \) under investigation. Let \( y \) denote the toxicity outcome which is either binary or quasi-binary (e.g., DLT or ETS) or continuous (e.g., TTB, TBS or TTP). For the motivating trial, after an appropriate transformation, we take the AUC as a continuous end point and model it by a normal distribution. [14] adopted the binomial and the normal distributions for binary (or quasi-binary) and continuous endpoints, respectively. Define \( \mu = E(y) \) and \( \mu_j = E(y|d_j) \).

Given the dose \( d_j \), the distribution of \( y \) belongs to the exponential family,

\[
f(y|d_j) = h(y) \exp \left\{ \eta(\theta_j) T(y) - A(\theta_j) \right\},
\]

where,

- \( \theta_j = \mu_j, \eta(\theta_j) = \log \left( \mu_j / (1 - \mu_j) \right) \),
- \( A(\theta_j) = -\log(1 - \mu_j), T(y) = y, \) and \( h(y) = 1 \), if \( y \) follows a binomial distribution;
- \( \theta_j = (\mu_j, \sigma^2), \eta(\theta_j) = \mu_j / \sigma^2, A(\theta_j) = \mu_j^2 / (2\sigma^2), T(y) = y, \) and \( h(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -y^2 / (2\sigma^2) \right\}, \) if \( y \) follows a normal distribution.

Let \( \phi_0 \) denote the target value of \( \mu \) for dose finding. Specifically, for binary or quasi-binary toxicity endpoints, \( \phi_0 \) is the target DLT probability; for continuous endpoints, \( \phi_0 \) is the targeted value of the TTB, TBS or TTP. Assume there are \( n_j \) patients treated at dose level \( d_j \) and let \( D_j = (y_1, \ldots, y_{n_j}) \) denote the observed toxicity data. Based on \( D_j \), the sample mean can be obtained as \( \hat{\mu}_j = \sum_{i=1}^{n_j} y_i / n_j \). For the interval-based design, dose transition decisions are made by comparing \( \hat{\mu}_j \) with the decision boundaries, \( \lambda_c(d_j, n_j, \phi_0) \) and \( \lambda_d(d_j, n_j, \phi_0) \). Specifically, if \( \hat{\mu}_j < \lambda_c(d_j, n_j, \phi_0) \), escalate to the higher dose level \( j + 1 \), and if \( \hat{\mu}_j > \lambda_d(d_j, n_j, \phi_0) \), de-escalate to the lower dose level \( j - 1 \), otherwise retain the same dose level \( j \). The selection of the decision boundaries \( \lambda_c(d_j, n_j, \phi_0) \) and \( \lambda_d(d_j, n_j, \phi_0) \) is critical because these two parameters essentially determine operating characteristics of a design.

Let the decisions retainment, escalation and de-escalation (each based on the current dose level), denoted as \( R, E \) and \( D \), respectively and let \( \overline{R} \) denote the decisions that are complementary to \( R \) (i.e., \( \overline{R} \) includes \( E \) and \( D \) ), and \( \overline{E} \) and \( \overline{D} \) denote the decisions that are complementary to \( E \) and \( D \), respectively. Following the same rule of [9], to obtain optimal decision boundaries under some criteria, the gBOIN [14] considers three point hypotheses \( H_0 : \mu_j = \phi_0, \quad H_1 : \mu_j = \phi_1, \quad H_2 : \mu_j = \phi_2 \) and minimize an incorrect decision probability \( \alpha \),

\[
\alpha = P(H_0)P(\overline{R}|H_0) + P(H_1)P(\overline{E}|H_1) + P(H_2)P(\overline{D}|H_2),
\]

where \( \phi_1 \) is a value deemed subtherapeutic such that dose escalation is warranted, and \( \phi_2 \) is a value deemed overly toxic such that dose de-escalation is required. Note that, \( H_0 \) indicates that the current dose is the MTD and we should retain the current dose for the next cohort of patient; \( H_1 \) indicates that the current dose is below the MTD and we should escalate the dose; and \( H_2 \) indicates that the current dose is overly toxic and we should deescalate the dose. Thus, the correct decisions under hypotheses \( H_0, H_1 \) and \( H_2 \) are retainment, escalation and de-escalation. Correspondingly, the incorrect decisions
under $H_0$, $H_1$ and $H_2$ are $\mathcal{R}$, $\mathcal{E}$ and $\mathcal{D}$, respectively. For example, under $H_0$ (i.e., the current dose is the target), the correct decision is to retain the current dose (i.e., $\mathcal{R}$), and incorrect decisions are dose escalation and de-escalation (i.e., $\mathcal{E}$ and $\mathcal{D}$). Taking a noninformative prior, i.e., $P(H_0) = P(H_1) = P(H_2) = 1/3$, and minimizing the incorrect decision probability $\alpha$ in Eq. (2), the decision boundaries can be obtained as (details can be found in [14]),

$$\lambda^*_e = \frac{A(\bar{\theta}_1) - A(\bar{\theta}_0)}{\eta(\bar{\theta}_1) - \eta(\bar{\theta}_0)}, \quad \lambda^*_d = \frac{A(\bar{\theta}_2) - A(\bar{\theta}_0)}{\eta(\bar{\theta}_2) - \eta(\bar{\theta}_0)}. \quad (3)$$

Specifically, when $y$ follows a Bernoulli or quasi-Bernoulli distribution, we have $\bar{\theta}_k = \varphi_k$, $A(\bar{\theta}_k) = -\log(1 - \varphi_k)$, $\eta(\bar{\theta}_k) = \log \{\varphi_k/(1 - \varphi_k)\}$. Then,

$$\lambda^*_e = \frac{\log \varphi_1 - \log \varphi_0}{\log \{1 - \varphi_0\}/\log \{1 - \varphi_1\}}, \quad \lambda^*_d = \frac{-\log \varphi_2}{\log \{1 - \varphi_0\}/\log \{1 - \varphi_2\}}, \quad (4)$$

which are exactly the same as boundaries provided by the original BOIN design [9]. When $y$ follows a normal distribution, we have $\bar{\theta}_k = (\varphi_k, \sigma_k^2)$, $A(\bar{\theta}_k) = \varphi_k^2 / (2\sigma_k^2)$, $\eta(\bar{\theta}_k) = \varphi_k/\sigma_k^2$. Then,

$$\lambda^*_e = \frac{\varphi_0 + \varphi_1}{2}, \quad \lambda^*_d = \frac{\varphi_0 + \varphi_2}{2}. \quad (5)$$

Based on the above decision boundaries, the gBOIN design is summarized as follows:

(a) Patients in the first cohort are treated at the lowest dose level or at a prespecified dose level.

(b) At the current dose level $j$, assign a dose to the next cohort of patients,

- if $\tilde{\mu}_j \leq \lambda^*_e$, escalate dose level to $j + 1$,
- if $\tilde{\mu}_j \geq \lambda^*_d$, de-escalate dose level to $j - 1$, and
- otherwise, i.e., $\lambda^*_e < \tilde{\mu}_j < \lambda^*_d$, retain the same dose level, $j$.

(c) This process is continued until the maximum sample size is reached or the trial is terminated because of excessive toxicities.

It is remarkable that the optimal decision boundaries ($\lambda^*_e, \lambda^*_d$) are free of $d_j$ and $n_j$, which means that the same pair of boundaries are used throughout the trial no matter which dose is the current dose, nor how many patients have been treated at the current dose.

**Adaptive gBOIN design**

Extensive simulation studies have shown that the gBOIN is transparent and simple to implement, and it yields good performance that is comparable or superior to more complicated model-based designs. As described in the “Introduction” section, the un-fixed boundaries may allow a flexibility to penalize mis-allocation rate of patients at over-toxic doses. To account for un-fixed boundaries, firstly, we reformulate the above three hypotheses as follows,

$$H_0 : \mu_j = \phi_0 \quad \text{versus} \quad H_1 : \mu_j = \phi_1,$$

and

$$H_0 : \mu_j = \phi_0 \quad \text{versus} \quad H_2 : \mu_j = \phi_2.$$

In the Bayesian paradigm, the Bayes factor in favor of the alternative hypothesis $H_1$ against a fixed null hypothesis $H_0$ is defined as,

$$BF_{10}(D_j) = \frac{P(H_1|D_j)/P(H_0|D_j)}{P(H_1)/P(H_0)}, \quad (6)$$

and the null hypothesis $H_0$ is rejected if $BF_{10}(D_j)$ exceeds a prespecified threshold $\gamma_1$. Similarly, the Bayes factor in favor of the alternative hypothesis $H_2$ against a fixed null hypothesis $H_0$ is defined as,

$$BF_{20}(D_j) = \frac{P(H_2|D_j)/P(H_0|D_j)}{P(H_2)/P(H_0)}, \quad (7)$$

and the null hypothesis $H_0$ is rejected if $BF_{20}(D_j)$ exceeds a prespecified threshold $\gamma_2$. Note that, if we want to put more penalties on over-toxic allocation, values of $\gamma_1$ and $\gamma_2$ would be different and presumably $\gamma_1$ should be greater than $\gamma_2$ since smaller $\gamma_2$ means decisions of de-escalation are easier made if over-toxicities occur. Given the prior odds $P(H_k)/P(H_0) = 1$ and the threshold $\gamma_k, (k = 1, 2)$, we can determine an alternative hypothesis that maximize the probability that the Bayes factor forms a test exceed the specified threshold $\gamma_k$. In other words, here we can choose the value of $\phi^*_k, (k = 1, 2)$ (this notation has been introduced in the “Introduction” section) to maximize $P(BF_{k0}(D_j) > \gamma_k)$.

By the Lemma 1 of [21], $\phi^*_1$ and $\phi^*_2$ can be obtained by

$$\phi^*_1 = \arg\max_{\mu_j < \phi_0} g_{\gamma_1}(\mu_j, \phi_0) \quad \text{and} \quad \phi^*_2 = \arg\min_{\mu_j > \phi_0} g_{\gamma_2}(\mu_j, \phi_0) \quad (8)$$

respectively, where $g_{\gamma_k}(\mu_j, \phi_0) = \frac{\log(\gamma_k) + n_j \{A(\bar{\theta}_j) - A(\bar{\theta}_0)\}}{\eta(\bar{\theta}_j) - \eta(\bar{\theta}_0)}, k = 1, 2$.

Specifically, for binomial distribution, $\phi^*_1$ and $\phi^*_2$ can be given as,

$$\phi^*_1 = \arg\max_{\mu_j < \phi_0} \log(\gamma_1) - n_j \left\{\log(1 - \mu_j) - \log(1 - \phi_0)\right\},$$

$$\phi^*_2 = \arg\min_{\mu_j > \phi_0} \log(\gamma_2) - n_j \left\{\log(1 - \mu_j) - \log(1 - \phi_0)\right\}. \quad (9)$$

Obviously, the values of $\phi^*_1$ and $\phi^*_2$ depend on the target $\phi_0$, the sample size $n_j$ and the threshold $\gamma_k, k = 1, 2$. Although their close forms cannot be obtained, they
can be solved via numerical optimization methods. For normal distribution, $\phi_1^*$ and $\phi_2^*$ can be given as,

\[
\phi_1^* = \phi_0 - \sigma \sqrt{\frac{2 \log \gamma_1}{n_j}},
\]
\[
\phi_2^* = \phi_0 + \sigma \sqrt{\frac{2 \log \gamma_2}{n_j}}.
\]  

(10)

Note that, for the normal distribution, values of $\phi_1^*$ and $\phi_2^*$ depend on the value of $\sigma$. So, if $\sigma$ is unknown, we can replace it with its sample estimation $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n_j} (y_i - \hat{\mu}_j)^2}{n_j}}$, or alternatively, we can take an inverse Gamma distribution with shape parameter $\alpha_0$ and rate parameter $\beta_0$ as its prior, then $\sigma$ can be replaced by using its posterior mean $\left(2\beta_0 + \sum_{i=1}^{n_j} (y_i - \hat{\mu}_j)^2\right) / (n_j + \alpha_0)$ with $\mu$ replaced by $\hat{\mu}_j$.

Replacing $\phi_1^*$ in $\lambda_e^*(k)$ and $\lambda_d^*(k)$ with $\phi_k^*$, $k = 1, 2$, we can get the adaptive shrinkage decision boundaries $\lambda_e^*(n_j)$ and $\lambda_d^*(n_j)$. Note that, for a standard binary toxicity endpoint, if we take the same values for $\gamma_k$, $k = 1, 2$, the boundaries are the same as the UMPI design [22]. Based on Lemma 2 in [21], we have the following double-shrinkage property theorem about the shrinkage boundaries $\lambda_e^*(n_j)$ and $\lambda_d^*(n_j)$.

**Theorem 1** As $n_j \rightarrow \infty$, the decision boundaries $\lambda_e^*(n_j)$ and $\lambda_d^*(n_j)$ will converge to the target $\phi_0$ at the rate of $O\left(\sqrt{\log \gamma_1}/n_j\right)$ and $O\left(\sqrt{\log \gamma_2}/n_j\right)$ respectively.

Theorem 1 introduces a double-shrinkage property for the proposed adaptive gBOIN design: The optimal values $\phi_k^*$ shrink toward the target toxicity probability $\phi_0$, and the optimal boundaries $\lambda_e^*(n_j)$ and $\lambda_d^*(n_j)$ based on each combination of $\phi_k^*$ and $\phi_k^*$ shrinkage toward the target value $\phi_0$.

Now we give the procedure of the proposed gBOINS design as follows.

(a) Patients in the first cohort are treated at the lowest dose level or at a prespecified dose level.

(b) At the current dose level $j$, to assign a dose to the next cohort of patients,

- if $\hat{\mu}_j \leq \lambda_d^*(n_j)$, escalate the dose level to $j + 1$,
- if $\hat{\mu}_j \geq \lambda_e^*(n_j)$, de-escalate the dose level to $j - 1$,
- otherwise, i.e., $\lambda_d^*(n_j) < \hat{\mu}_j < \lambda_e^*(n_j)$, retain the same dose level, $j$.

(c) This process is continued until the maximum sample size is reached or the trial is terminated because of excessive toxicities.

After the trial has been completed, we use the pooled adjacent violators algorithm [23] to select a dose level as the MTD. Denote the isotonicity transformed values of the observed value $(\hat{\mu}_j)$ by $(\hat{\mu}_j)$, to be specific, for finding the MTD, we select dose $\lambda^*$, for which the isotonic estimate of the toxicity rate $\hat{\mu}_{\lambda^*}$ is closest to $\phi_0$; if there are ties for $\hat{\mu}_{\lambda^*}$, we select from the ties the highest dose level when $\hat{\mu}_{\lambda^*} < \phi_0$ or the lowest dose level when $\hat{\mu}_{\lambda^*} > \phi_0$.

For patient safety, we impose the following overdose control rule when using the gBOIN design.

If $P(\mu_j > \phi_0 \mid D_j) > 0.95$ 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.

Posterior probability $P(\mu_j > \phi_0 \mid D_j) > 0.95$ can be evaluated on the basis of a beta-binomial model for the binary or quasi-binary endpoint, assuming $\mu_j$ follows a vague beta prior, e.g., $\mu_j \sim \text{beta}(1, 1)$. For normal endpoint $y$ with mean $\mu_j$ and variance $\sigma^2_j$, assuming noninformative prior $(\mu, \sigma^2) \propto \sigma^{-2}$, the posterior distribution of $\mu_j$ follows a $t$ distribution with $n_j - 1$ degrees of freedom, mean $\hat{\mu}_j$ and scale $n_j^{-1} \sum_{i=1}^{n_j} (y_i - \hat{\mu}_j)^2$.

**Design properties**

From a practical viewpoint, a natural requirement for dose-finding trials is that dose escalation should be not allowed if the observed toxicity rate or mean toxicity score at the current dose is higher than the target, and dose de-escalation should not be allowed if the observed toxicity rate or mean toxicity score at the current dose is lower than the target. [9] referred to this finite sample property as "long-term memory coherent", which is an extension of a similar concept originally proposed by [24]. That original definition of design coherence requires the prohibition of dose escalation (or de-escalation) 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 without considering responses from patients who were previously enrolled and treated, [9] refers this definition as "short-term memory coherence". Clearly, short-term memory coherence is a stronger counterpart than long-term memory coherence.

As shown in the Appendix, the gBOINS design has the following desirable finite-sample property.

**Theorem 2** The gBOINS design is long-term memory coherent in the sense that the design will never escalate the dose when $\hat{\mu}_j > \phi_0$; and will never de-escalate the dose when $\hat{\mu}_j < \phi_0$.

To further enhance safety of the design, we let the upper boundary $\phi_2^*$ have a little bit faster shrinking rate than that
of the lower boundary \( \phi^*_j \), since more strict or smaller \( \phi^*_j \) has less risk of exposing participated patients to over-toxic doses. We propose to take \( \gamma_k \) as \( \gamma_k = \exp(c_k n_j^e), k = 1, 2, 1 \leq \epsilon_1 \leq \epsilon_2 > 0 \) and \( 0 < c_1 < n_j^{1-\epsilon_1} \log(1/(1-\phi_0)) \) and \( c_1 < c_2 < n_j^{1-\epsilon_2} \log(1/\phi_0) \). It can be shown that the proposed adaptive gBOIN design has the following desirable large-sample property.

**Theorem 3** As the number of patients goes to infinity, the dose assignment and the selection of the MTD under the gBOINS design converge almost surely to dose level \( j^* \), if \( \mu_j = \phi_0 \).

According to Theorem 1, the condition \( \gamma_k = \exp(c_k n_j^e), c_k > 0, k = 1, 2 \), imposed here to leverage the converge rate of \( \lambda^*_j(n_j) \) and \( \lambda^*_j(n_j) \), yielding \( P[\hat{\mu}_j \in (\lambda^*_j(n_j), \lambda^*_j(n_j))] = 1 \), because \( \hat{\mu}_j \) converges in probability to \( \mu_j \) at the \( \sqrt{n} \) rate. Following the proof of Theorem 1 of [25], the result can be directly obtained and is omitted here.

**Practical implementation**

To implement the proposed gBOINS design in practice, we need to specify the values of \( \epsilon_k \) and \( c_k, k = 1, 2 \). We recommend the \( \epsilon_k = 0.5, k = 1, 2 \). The values of \( c_k, k = 1, 2 \) need to be calibrated by extensive simulation studies, and even there are no uniform values for different type of endpoints with the same target. For the normal endpoints, the shrinkage boundaries depend on the estimate of \( \sigma \), this will influence the pre-tabulation and the simplicity of gBOINS. For practical applications, we suggest to replace it with \( \lambda(n) \). Note that a big (small) value of \( \lambda^*_j(n_j) \) (or \( \lambda^*_j(n_j) \)) will make dose escalation (de-escalation) rapidly, this may lead serious safety problems and reduce the efficiency of the design when the sample size is small, since the smaller value of the sample size the bigger variance of \( \mu_j \). To avoid this adverse event problem and improve the design’s efficiency, in practice, we introduce a lead-in process in a trial to follow the original gBOIN design for a pre-specified number of patients (denoted as \( N_0 \)). After \( n_j > N_0 \), the trial is then switched to the gBOINS design. For our simulations, \( N_0 = 6 \) is recommended. Table 1 shows examples of the values of \( (\lambda^*_j(n_j), \lambda^*_j(n_j)) \) for target \( \phi_0 = 0.2 \) and \( \phi_0 = 0.3 \).

**Simulation**

**Toxicity as a binary endpoint**

We test the performance of the gBOINS design by comparing it to the gBOIN design under four different metrics: the percentage of correct selection (PCS) of the MTD, the average number of patients allocated to the MTD, the risk of overdosing, which is defined as the percentage of simulated trials in which a large percentage (e.g., more than 60% or 80%) of patients are treated at doses above the MTD and the risk of underdosing which is defined as the percentage of simulated trials in which more than 80% of patients are treated at doses below the MTD. We investigated two target toxicity rates \( \phi_0 = 0.2 \) and \( \phi_0 = 0.3 \), and for each of the target toxicity rate, we examined 16 representative toxicity scenarios with various parameters of \( \phi_0 = 0.2 \), \( c_1 = \log(1.05)/3 \) and \( c_2 = \log(1.05)/3 \), and \( c_1 = \log(1.1)/3 \) and \( c_2 = \log(1.1)/3 \) when \( \phi_0 = 0.3 \). Table 2 which were reproduced from [27]. All examined scenarios are varied in the location of the MTD and the gaps around the MTD. For each scenario, 30 patients and 10 cohorts were assumed.

Figures 1 and 2 present the results based on 4000 simulated trials. As shown in Fig. 1, when the target is 0.2, for all 16 scenarios the performance of gBOINS and gBOIN are comparable in the sense of percentage of correct selection of the target dose and the average number of patients allocated to the MTD. While the gBOIN has a higher risk of overdosing, for most scenarios, acrom scenarios 1 to

**Table 1** Dose escalation and de-escalation boundaries for Bernoulli and continuous toxicity endpoint, with \( \phi_0 = 0.2, \phi_0 = 0.3, \epsilon_k = 0.5, k = 1, 2 \) and \( N_0 = 6 \)

| End point | \( n_j \) | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 |
|-----------|---------|---|---|---|---|---|---|---|---|---|---|
| Bernoulli | \( \phi_0 = 0.2 \) | \( \lambda^*_j(n_j) \) | 0.16 | 0.16 | 0.16 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 |
| or quasi- | \( \phi_0 = 0.3 \) | \( \lambda^*_j(n_j) \) | 0.24 | 0.24 | 0.24 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 |
| Bernoulli | \( \phi_0 = 0.2 \) | \( \lambda^*_j(n_j) \) | 0.36 | 0.36 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.32 |
| Continuous | \( \phi_0 = 0.2 \) | \( \lambda^*_j(n_j) \) | 0.24 | 0.24 | 0.24 | 0.25 | 0.25 | 0.25 | 0.25 | 0.26 | 0.26 |
|           | \( \phi_0 = 0.3 \) | \( \lambda^*_j(n_j) \) | 0.36 | 0.36 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 |

For the binary endpoint, \( c_1 = \log(1.05) / 3 \) and \( c_2 = \log(1.05) / 3 \) for \( \phi_0 = 0.2 \), and \( c_1 = \log(1.1) / 3 \) and \( c_2 = \log(1.1) / 3 \) for both \( \phi_0 = 0.2 \) and \( \phi_0 = 0.3 \).
16. In addition, compared to the gBOIN design, the proposed gBOINS allocated fewer patients to sub-therapeutic doses for most scenarios, which may be explained by a higher risk of underdosing 80% for the gBOIN. Figure 2 shows that, when the target is 0.3, the performance of gBOINS and gBOIN are comparable, and the proposed gBOINS has a lower risk of overdosing.

At the end of the “Toxicity as a binary endpoint” section, we also conducted simulation studies to investigate the performance of the gBOINS with respect to different sample sizes. We consider four scenarios in Table 3 and the simulation results based on 4000 replications are presented in Fig. 3. As shown in the first two pictures on the left panel of Fig. 3, for scenarios 1 and 3, there was...
only one dose lying inside the interval (0.16, 0.24) and (0.24, 0.36) respectively, the performance of gBOINS is comparable to the gBOIN design. For scenarios 2 and 4, the simulation results are depicted on the right panel of Fig. 3, there were two doses lying inside the intervals and the gBOINS outperformed the gBOIN when the sample size was greater than 90.

**Toxicity as a quasi-binary endpoint**

We evaluated the performance of the gBOINS design when the toxicity endpoint was a quasi-binary endpoint (e.g., ETS) by comparing it to the gBOIN design method based on the ten scenarios considered by [14](see Table 4). Following [18], we adopted the following ETS definition: grades 0 and 1 were of no concern (no DLT); a grade-2 toxicity was equivalent to 0.5 DLT; a grade-3 toxicity counted as one DLT; and a grade-4 toxicity was equivalent to 1.5 DLT. The target ETS was 0.47, derived from the target profile of 49% grade 0 and grade 1, 18% grade 2, 23% grade 3, and 10% grade 4. That is, the target ETS = \[ \frac{0.49 \times 0 + 0.18 \times 0.5 + 0.23 \times 1.0 + 0.10 \times 1.5}{1} = 0.47. \]

A total sample of 30 patients in 10 cohorts was used in the

| Scenario | Target DLT rate = 20% |  | Target DLT rate = 30% |  |
|----------|-----------------------|---|-----------------------|---|
| 1        | 0.01                  | 0.10 | 0.20                 | 0.30 | 0.35 |
| 2        | 0.01                  | 0.18 | 0.20                 | 0.30 | 0.35 |
| 3        | 0.01                  | 0.20 | 0.30                 | 0.40 | 0.50 |
| 4        | 0.01                  | 0.25 | 0.30                 | 0.40 | 0.50 |

**Table 3** Four true toxicity scenarios with the target DLT rates of 0.2 and 0.3 in boldface
Fig. 3 The relationships of operating characteristics and sample sizes of gBOIN and gBOINS when the target toxicity rates are 20% and 30%. The two pictures on the left panel, only one dose lying inside (0.16, 0.24) and (0.24, 0.36) respectively for the top and bottom panel; on the right panel, there are two doses lying inside their corresponding intervals.

Fig. 3 reveals the relationship between sample size and correct selection rate for gBOIN and gBOINS when the target DLT rate is 20% and 30%. The graphs show that as the sample size increases, the correct selection rate also increases for both designs. The two panels on the left illustrate the cases where only one dose lies within the specified interval, while the right panels show the scenarios with two doses within the intervals.

In general, the performance of gBOINS is comparable with the gBOIN design in terms of PCS, number of patients allocated to the MTD, while the gBOIN assigns more patients to the overly toxic doses above the MTD. Table 5 shows the results based on 4,000 simulated trials. For Scenarios 1, 2, 3, 4, 5, and 8, the performance of gBOINS is comparable with the gBOIN design in terms of PCS, number of patients allocating to the MTD, while the gBOIN assigns more patients to the overly toxic doses above the MTD. For example, in scenario 3, in which dose level 2 was the MTD, the gBOINS design yielded a PCS of 48% and allocated 14.5 patients to the MTD; the gBOIN yielded a PCS of 47% and allocated 12.1 patients to the MTD. While the gBOINS assigned 2.7 fewer patients than the gBOIN design to the overly toxic doses. In scenario 6, in which dose level 1 was the MTD, the PCS of the gBOINS was 77% and has a 8% higher than that of the gBOIN design. In scenario 7, the MTD was located at the low level 2 and the gBOIN design yielded a of 55%, which was 4% higher than that of the gBOINS. While the gBOINS allocated more patients to the MTD and assigned fewer patients to the overly toxic doses. In scenario 9, in which dose level 2 was the MTD, the gBOINS design yielded a PCS of 96% and was 5% higher than that of the gBOIN design. The number of patients allocated to MTD was similar, while the gBOINS assigned 3.4 fewer patients than the gBOIN design to the overly toxic doses. In scenario 10, in which dose level 4 was the MTD, the gBOINS design yielded a PCS of 76% and was 5% higher than that of the gBOIN design. The gBOIN design assigned 13.3 patients to MTD and was
higher than that of the gBOINS, while gBOINS assigned fewer patients to the overly toxic doses. In addition, the gBOINS design yielded a 6% chose the overly toxic doses as MTD and had a substantially improvement compared with the gBOIN design which has a 26% chose the overly toxic doses as the MTD.

**Continuous outcomes**

In this section, we consider ten scenarios with continuous outcomes in Table 6, all responses follow the normal distribution adopted by [20]. For the first six scenarios, the response $y$ at the dose level $x \in \{1, 2, 3, 4, 5, 6\}$ follows a normal distribution $N(0.05 + 0.05x, 0.05^2x^2)$ and when the target at $x = 1$, a sample size of 15 was used and when the target at other dose levels, a sample size of 60 was used. Cohort size of 1 was used for all scenarios. For the rest scenarios, scenarios 7 to 10, the response $y$ at the dose $x \in \{1, 2, 3, 4, 5, 6\}$ also followed the normal distribution $N(0.05 + 0.05x, 0.05^2x^2)$, and a moderate large sample size of 100 was used. And the $c_k$, $k = 1, 2$, are set to be $c_1 = \log(1.1)/3$ and $c_2 = \log(1.1)$ throughout this subsection.

Table 6 shows that when the sample size is 15 for scenario 1 and 60 for scenarios 2 to 6, performance of the gBOINS design is comparable with the gBOIN design, in correct selection percentage and number of patients treated at the target dose. While for the last four scenarios, when the sample size is moderate large, the gBOINS outperformed the gBOIN design in correct selection percentage and was comparable with gBOIN design in number of patients allocated to the target dose. Specifically, in scenario 7, in which dose level 3 was the target dose, gBOINS yielded a PCS of 89% and allocated 61.7 patients to the MTD, whereas the gBOIN yield a PCS of 84% and allocated 64 patients to the MTD. In scenario 8, the MTD was located at the dose level 4, the PCS of gBOINS was 75% and 5% higher than that of the gBOIN. In scenarios 9-10, we also see that the PCS of gBOINS was superior to the gBOIN in PCS and was comparable with the gBOIN in number of patients allocated to the MTD.

| Grade | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Scenario 1 | 0.83 | 0.75 | 0.62 | 0.51 | 0.34 | 0.19 | 0.92 | 0.85 | 0.70 | 0.55 | 0.24 | 0.00 |
| Scenario 2 | 0.12 | 0.15 | 0.18 | 0.19 | 0.16 | 0.11 | 0.03 | 0.05 | 0.10 | 0.15 | 0.26 | 0.36 |
| Scenario 3 | 0.04 | 0.07 | 0.11 | 0.14 | 0.15 | 0.11 | 0.03 | 0.07 | 0.14 | 0.21 | 0.35 | 0.49 |
| Scenario 4 | 0.01 | 0.03 | 0.09 | 0.16 | 0.35 | 0.59 | 0.02 | 0.03 | 0.06 | 0.09 | 0.15 | 0.21 |
| Scenario 5 | 0.78 | 0.56 | 0.50 | 0.40 | 0.30 | 0.16 | 0.88 | 0.64 | 0.52 | 0.35 | 0.17 | 0.00 |
| Scenario 6 | 0.14 | 0.19 | 0.18 | 0.17 | 0.15 | 0.09 | 0.04 | 0.12 | 0.16 | 0.22 | 0.28 | 0.39 |
| Scenario 7 | 0.06 | 0.12 | 0.14 | 0.15 | 0.14 | 0.10 | 0.06 | 0.17 | 0.22 | 0.30 | 0.38 | 0.52 |
| Scenario 8 | 0.02 | 0.12 | 0.18 | 0.28 | 0.41 | 0.65 | 0.02 | 0.07 | 0.10 | 0.13 | 0.17 | 0.23 |
| Scenario 9 | 1.00 | 0.91 | 0.88 | 0.86 | 0.80 | 0.65 | 0.50 | 0.38 | 0.29 | 0.19 | 0.13 | 0.08 |
| Scenario 10 | 0.14 | 0.18 | 0.17 | 0.15 | 0.09 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Scenario 11 | 0.08 | 0.24 | 0.32 | 0.43 | 0.55 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Scenario 12 | 1.00 | 0.91 | 0.88 | 0.86 | 0.80 | 0.65 | 0.50 | 0.38 | 0.29 | 0.19 | 0.13 | 0.08 |
| Scenario 13 | 1.00 | 0.91 | 0.88 | 0.86 | 0.80 | 0.65 | 0.50 | 0.38 | 0.29 | 0.19 | 0.13 | 0.08 |
| Scenario 14 | 1.00 | 0.91 | 0.88 | 0.86 | 0.80 | 0.65 | 0.50 | 0.38 | 0.29 | 0.19 | 0.13 | 0.08 |
| Scenario 15 | 1.00 | 0.91 | 0.88 | 0.86 | 0.80 | 0.65 | 0.50 | 0.38 | 0.29 | 0.19 | 0.13 | 0.08 |
Table 5 Simulation results comparing the gBOINS design with the gBOIN when the equivalent toxicity score (ETS) is used as the quasi-binary toxicity endpoint. The target dose (i.e., MTD) is in boldface.

| Selection percentage | Average number of patients treated |
|----------------------|-----------------------------------|
| d1   | d2   | d3   | d4   | d5   | d6   | d1   | d2   | d3   | d4   | d5   | d6   |
|-------|------|------|------|------|------|------|------|------|------|------|------|
| **Scenario 1**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.12 | 0.19 | 0.34 | **0.48** | 0.76 | 1.05 | 3.5  | 5.1  | 8.4  | **8.9** | 3.7  | 0.4  |
| gBOIN 0.00 | 0.02 | 0.29 | **0.56** | 0.12 | 0.01 | 4.7  | 6.8  | 10.2 | **7.1** | 1.3  | 0.0  |
| gBOINS 0.00 | 0.03 | 0.32 | 0.53  | 0.12 | 0.00 | 3.8  | 5.0  | 8.7  | 9.5  | 3.0  | 0.1  |
| **Scenario 2**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.08 | 0.14 | 0.28 | **0.42** | 0.70 | 0.98 | 3.5  | 4.4  | 7.5  | **9.9** | 4.3  | 0.4  |
| gBOIN 0.00 | 0.00 | 0.16 | **0.66** | 0.18 | 0.00 | 3.8  | 5.0  | 8.7  | 9.5  | 3.0  | 0.1  |
| gBOINS 0.00 | 0.01 | 0.14 | 0.63  | 0.22 | 0.00 | 3.8  | 5.0  | 8.7  | 9.5  | 3.0  | 0.1  |
| **Scenario 3**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.16 | **0.40** | 0.50 | 0.66 | 0.83 | 0.12 | 6.3  | **12.1** | 7.7  | 3.2  | 0.5  | 0.0  |
| gBOIN 0.02 | **0.47** | 0.36 | 0.13 | 0.00 | 0.00 | 6.8  | 14.5 | 7.0  | 1.6  | 0.1  | 0.0  |
| gBOINS 0.06 | 0.48 | 0.36 | 0.09 | 0.01 | 0.00 | 6.8  | 14.5 | 7.0  | 1.6  | 0.1  | 0.0  |
| **Scenario 4**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.11 | 0.34 | **0.45** | 0.60 | 0.78 | 1.06 | 4.8  | 9.0  | **10.8** | 4.4  | 0.9  | 0.1  |
| gBOIN 0.01 | 0.28 | **0.47** | 0.23 | 0.01 | 0.00 | 5.0  | 11.8 | 10.0 | 3.0  | 0.2  | 0.0  |
| gBOINS 0.01 | 0.27 | 0.50 | 0.21 | 0.01 | 0.00 | 5.0  | 11.8 | 10.0 | 3.0  | 0.2  | 0.0  |
| **Scenario 5**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.00 | 0.06 | 0.09 | 0.10 | 0.16 | **0.32** | 3.0  | 3.1  | 3.3  | 3.4  | 4.6  | **12.6** |
| gBOIN 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | **0.96** | 3.0  | 3.8  | 4.2  | 4.3  | 5.2  | 9.4  |
| gBOINS 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.94  | 3.0  | 3.8  | 4.2  | 4.3  | 5.2  | 9.4  |
| **Scenario 6**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.44 | 0.63 | 0.80 | 1.01 | 1.16 | 1.29 | 19.4 | 8.0  | 1.7  | 0.2  | 0.0  | 0.0  |
| gBOIN 0.69 | 0.29 | 0.01 | 0.00 | 0.00 | 0.00 | **19.4** | 8.0  | 1.7  | 0.2  | 0.0  | 0.0  |
| gBOINS 0.77 | 0.21 | 0.02 | 0.00 | 0.00 | 0.00 | 22.9 | 6.5  | 0.7  | 0.0  | 0.0  | 0.0  |
| **Scenario 7**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.19 | **0.45** | 0.57 | 0.73 | 0.9  | 1.17 | 8.9  | **13.0** | 6.2  | 1.7  | 0.2  | 0.0  |
| gBOIN 0.11 | **0.55** | 0.28 | 0.06 | 0.00 | 0.00 | 8.8  | 14.6 | 5.6  | 0.9  | 0.0  | 0.0  |
| gBOINS 0.15 | 0.51 | 0.30 | 0.04 | 0.00 | 0.00 | 8.8  | 14.6 | 5.6  | 0.9  | 0.0  | 0.0  |
| **Scenario 8**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.08 | 0.24 | 0.32 | **0.43** | 0.55 | 0.75 | 2.7  | 5.6  | 7.2  | **8.5** | 4.4  | 1.6  |
| gBOIN 0.00 | 0.06 | 0.22 | **0.41** | 0.28 | 0.04 | 3.9  | 7.8  | 9.3  | 6.7  | 2.1  | 0.2  |
| gBOINS 0.00 | 0.05 | 0.28 | 0.42  | 0.23 | 0.03 | 3.9  | 7.8  | 9.3  | 6.7  | 2.1  | 0.2  |
| **Scenario 9**     |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.19 | **0.45** | 0.57 | 0.73 | 0.9  | 1.17 | 3.2  | **22.7** | 4.1  | 0.0  | 0.0  | 0.0  |
| gBOIN 0.00 | **0.91** | 0.09 | 0.00 | 0.00 | 0.00 | 5.6  | 23.7 | 0.7  | 0.0  | 0.0  | 0.0  |
| gBOINS 0.00 | 0.96 | 0.04 | 0.00 | 0.00 | 0.00 | 5.6  | 23.7 | 0.7  | 0.0  | 0.0  | 0.0  |
| **Scenario 10**    |      |      |      |      |      |      |      |      |      |      |      |
| ETS 0.08 | 0.24 | 0.32 | **0.43** | 0.55 | 0.75 | 3.0  | 3.5  | 5.0  | **13.3** | 3.8  | 1.3  |
| gBOIN 0.00 | 0.00 | 0.03 | **0.71** | 0.23 | 0.03 | 4.3  | 7.0  | 10.4 | 8.1  | 0.2  | 0.0  |
| gBOINS 0.00 | 0.00 | 0.19 | 0.76  | 0.05 | 0.01 | 4.3  | 7.0  | 10.4 | 8.1  | 0.2  | 0.0  |
Table 6: Simulation results comparing gBOINS with gBOIN, when continuous toxicity endpoint is used. The target, correct selection percentage and correct allocation number for each scenario are in boldface.

| Scenario | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | Allocation $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ |
|----------|-------|-------|-------|-------|-------|-------|----------------|-------|-------|-------|-------|-------|
| Scenario 1 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.93           | 0.06  | 0.01  | 0.00  | 0.00  | 0.00  |
| Scenario 2 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.03           | 0.91  | 0.06  | 0.00  | 0.00  | 0.00  |
| Scenario 3 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.02           | 0.92  | 0.06  | 0.00  | 0.00  | 0.00  |
| Scenario 4 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.10           | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  |
| Scenario 5 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.00           | 0.09  | 0.78  | 0.12  | 0.01  | 0.00  |
| Scenario 6 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.00           | 0.00  | 0.23  | 0.60  | 0.16  | 0.02  |
| Scenario 7 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.00           | 0.00  | 0.00  | 0.28  | 0.53  | 0.19  |
| Scenario 8 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.00           | 0.00  | 0.00  | 0.03  | 0.35  | 0.61  |
| Scenario 9 | 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35  | 0.00           | 0.00  | 0.17  | 0.69  | 0.13  | 0.01  |
| Scenario 10| 0.10  | 0.15  | 0.20  | 0.25  | 0.30  | 0.35 | 0.00           | 0.00  | 0.00  | 0.26  | 0.57  | 0.18  |

Conclusion

We proposed a new phase I trial design that incorporates toxicity grades into the dose finding trials. The proposed gBOINS design unifies the continuous and quasi-binary toxicity endpoints as well as the standard binary endpoint. Different from the gBOIN design, the decision boundaries of gBOINS design, $\lambda^*_e(n_j)$ and $\lambda^*_d(n_j)$ were adaptive, which provides the statistician a flexible tool to control the over-toxicities. The design can also converge to the target toxicity probability as the sample size goes to infinity. This unique convergence property of gBOINS was demonstrated both theoretically and numerically. Compared to the gBOIN design, when there were more than one doses lying inside the decision boundaries $(\lambda^*_e, \lambda^*_d)$ determined by the gBOIN design, the gBOINS had a substantial improvement in terms of the PCS when the sample size was moderate large. Also, we showed that when the sample size was small, the performance of gBOINS design was comparable with the gBOIN design in terms of the PCS and can allocate more patients to safe doses by simulations.

Although the proposed gBOINS design focus on phase I trial designs, similarly to the BOIN-ET proposed by [28], it can be directly extended to the phase I/II designs. One limitation of the gBOINS design is that it assumes toxicity outcome can be observed quickly enough to make the
dose assignment decisions for each enrolled cohort. One approach to extend the gBOINS design to accommodate late-onset or delayed outcomes, for example, would be to use the Bayesian data augmentation approach [29, 30] or the approximated likelihood approach [31]. This is a topic of our future research.

Acknowledgments
The research is supported by the National Natural Science Foundation of China (grant 11901519, grant 12001378), China Postdoctoral Science Foundation (2019M661416), Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A1515110449) and Natural Science Foundation of Guangdong Province (Grant No. 2020A1515010372). Pan’s research is partially supported by American Lebanese Syrian Associated Charities (ALSAC).

Funding
The research is supported by the National Natural Science Foundation of China (grant 11901519, grant 12001378), China Postdoctoral Science Foundation (2019M661416), Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A1515110449) and Natural Science Foundation of Guangdong Province (Grant No. 2020A1515010372). Pan’s research is partially supported by American Lebanese Syrian Associated Charities (ALSAC).

Authors’ contributions
HP and RM contributed to the conception, design, and development of the R code of this work. ZG and GX reviewed and approved the accompanying R code. Both HP and RM wrote, revised, and approved the final manuscript.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
1Clinical Research Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. 2College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China. 3Jiangsu Hengrui Medicine Co, Ltd, Shanghai 201203, China. 4Department of Biostatistics, St. Jude Children’s Research Hospital, 38105 Memphis, TN, USA.

Received: 10 March 2021 Accepted: 19 October 2021
Published online: 13 December 2021

References
1. Storer BE. Design and analysis of phase I clinical trials. Biometrics. 1989;45: 925–37.
2. Simon R, Rubinstein L, Arbuck SG, Christian MC, Freidlin B, Collins J. Accelerated titration designs for phase I clinical trials in oncology. J Natl Cancer Inst. 1997;89:1138–47.
3. Durham SD, Flourney N, Rosenberger WF. A random walk rule for phase I clinical trials. Biometrics. 1997;53:745–60.
4. 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.
5. Cheung YK. Dose Finding by the Continual Reassessment Method. New York: CRC Press; 2011.
6. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. Stat Med. 1998;17:1103–20.
7. Tighiouart M, Rogatko A. Dose finding with escalation with overdose control (EWOC) in cancer clinical trials. Stat Sci. 2010;25:217–26.
8. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. J Am Stat Assoc. 2009;104:954–68.
9. Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. J R Stat Soc Ser C. 2015;64:507–23.
10. Yan F, Mandrekar SJ, Yuan Y. Bayesian probability interval design approach for phase I clinical trials. Clin Cancer Res. 2012;17:3994–4003.
11. Brahmer JR, Drake CG, Wollner I, Powderly JD, Picus J, Shafman WH, Stankevich E, Pons A, Salay TM, McMillin TL, et al. Phase I study of single-agent anti–programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. J Clin Oncol. 2010;28:3167–75.
12. Le Tourneau C, Diéras V, Tresca P, Cacheux W, Paolelli X. Current challenges for the early clinical development of anticancer drugs in the era of molecularly targeted agents. Target Oncol. 2010;5:65–72.
13. Penel N, Adenis A, Clisant S, Bonnetere J. Nature and subjectivity of dose-limiting toxicities in contemporary phase I trials: comparison of cytotoxic versus non-cytotoxic drugs. Invest New Drugs. 2011;29:1414–9.
14. Mu R, Yuan Y, Xu J, Mandrekar SJ, Yin J. gBOIN: a unified model-assisted phase I trial design accounting for toxicity grades, and binary or continuous end points. J R Stat Soc Ser C. 2019;68:289–308.
15. Bekele BN, Thall PF. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. J Am Stat Assoc. 2000;95:26–35.
16. Lee S, Hershman D, Martin P, Leonard J, Cheung Y. Toxicity burden score: a novel approach to summarize multiple toxic effects. Ann Oncol. 2012;23:537–41.
17. Ezzelfani M, Zohar S, Qin R, Mandrekar SJ, Deley M-CL. Dose-finding designs using a novel quasi-continuous endpoint for multiple toxicities. Stat Med. 2013;32:2728–46.
18. Yuan Z, Chappell R, Bailey H. The continual reassessment method for multiple toxicity grades: A Bayesian quasi-likelihood approach. Biometrics. 2007;63:173–9.
19. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J Appl Econom. 1996;11:619–32.
20. Ivanova A, Kim SH. Dose finding for continuous and ordinal outcomes with a monotone objective function: A unified approach. Biometrics. 2009;65:307–15.
21. Johnson VE. Uniformly most powerful bayesian tests. Ann Stat. 2013;41:1716–41.
22. Lin R, Yin G. Uniformly most powerful bayesian interval design for phase I dose-finding trials. Pharm Stat. 2018;17:710–24.
23. Barlow RE, Bartholomew DJ, Bremner J, Brunk HD. Statistical Inference Under Order Restrictions. London: Wiley; 1972.
24. Cheung YK. Coherence principles in dose-finding studies. Biometrika. 2005;92:863–73.
25. Duan HP, Uma R, Diao X, Quan X. Dose-finding designs: an approach to optimal designs for dose-finding trials. Stat Med. 2021;40:3994–4003.
26. Yuan Y, Hess KR. Hilsenbeck SG. Gilbert MR. Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials. Clin Cancer Res. 2016;22:4291–301.
27. Yuan Y, Hess KR, Hilsenbeck SG, Gilbert MR. Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials. Clin Cancer Res. 2016;22:4291–301.
28. Takeda K, Taguri M, Morita S. Boin-et: Bayesian optimal interval design for dose finding based on both efficacy and toxicity outcomes. Pharm Stat. 2018;17:383–95.
29. Liu S, Yin G, Yuan Y. Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity. Ann App Stat. 2013;7:1837–2457.
30. Jin H, Liu S, Thall P, Yuan Y. Using data augmentation to facilitate conduct of phase I/Ii clinical trials with delayed outcomes. J Am Stat Assoc. 2014;109:525–36.
31. Lin R, Yuan Y. Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. Biostatistics. 2019;20:1807–24.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.