



Comparison Between Simultaneous and Sequential Utilization of Safety and Efficacy for Optimal Dose Determination in Bayesian Model-Assisted Designs

Ran Li¹ · Kentaro Takeda¹ · Alan Rong¹

Received: 21 November 2022 / Accepted: 21 March 2023 / Published online: 22 April 2023
© The Author(s), under exclusive licence to The Drug Information Association, Inc 2023

Abstract

It has become quite common in recent early oncology trials to include both the dose-finding and the dose-expansion parts within the same study. This shift can be viewed as a seamless way of conducting the trials to obtain information on safety and efficacy hence identifying an optimal dose (OD) rather than just the maximum tolerated dose (MTD). One approach is to conduct a dose-finding part based solely on toxicity outcomes, followed by a dose expansion part to evaluate efficacy outcomes. Another approach employs only the dose-finding part, where the dose-finding decisions are made utilizing both the efficacy and toxicity outcomes of those enrolled patients. In this paper, we compared the two approaches through simulation studies under various realistic settings. The percentage of correct ODs selection, the average number of patients allocated to the ODs, and the average trial duration are reported in choosing the appropriate designs for their early-stage dose-finding trials, including expansion cohorts.

Keywords Bayesian adaptive dose-finding design · Expansion cohort · Optimal dose · Seamless design · Sequential enrollment · Oncology

Introduction

With the emergence of immune-oncology therapy, the primary goal of dose-finding trials has been broadened from finding the maximum tolerated dose (MTD) to identifying the optimal dose (OD), defined as the tolerable dose having sufficient efficacy from a clinical point of view and several mathematical formulas has been proposed [1–3]. Unlike traditional cytotoxic agents, for immune-oncology therapy, dose–toxicity and dose–efficacy relationships are often unpredictable, where the efficacy may not monotonically increase with the dose level, and the OD may be well below the MTD. Therefore, there is a greater need to utilize efficacy and toxicity information to identify the OD. In 2018, the US Food and Drug Administration released draft guidance [4] recommending multiple expansion cohort trials to expedite oncology drug development. Several methods

are proposed to estimate the ODs from multiple expansion cohorts [5].

Two approaches are often used to determine the OD based on both the efficacy and toxicity outcomes in oncology early phase trials. The first approach is to conduct dose-finding using toxicity information to find the MTD, initiate multiple expansion cohorts, and then select the OD using efficacy outcomes from these expansion cohorts. For the dose-finding part, based solely on toxicity, lots of trial designs are proposed, for example, the algorithm-based designs such as the 3 + 3 design [6], the model-based designs such as the continual reassessment method [7] and the Bayesian logistic regression model [8], and the model-assisted designs such as the modified toxicity probability interval design [9], the Bayesian optimal interval (BOIN) design [10], and the keyboard design [11]. The model-assisted designs are getting popular due to the high performance in terms of correctly selecting the MTD and the simplicity. For the expansion part, Simon’s optimal two-stage design [12] is the most well-known design among all frequentist approaches. In addition, several Bayesian designs have been proposed. For example, Thall and Simon [13] have proposed a posterior probability approach to monitor phase II trials. Zhou et al.

✉ Ran Li
ran-li@astellas.com

¹ Biostatistics, Data Science, Astellas Pharma Inc, 1 Astellas Way, N3.272.A, Northbrook, IL 60062, USA

[14] have proposed the Bayesian optimal phase 2 (BOP2) design, which maximizes the power for detecting effective treatments while controlling the type I error. The second approach to finding the OD is simultaneously using the efficacy and toxicity outcomes to guide each cohort's dose assignment. The approach needs only a dose-finding part instead of dose-finding and dose-expansion parts. Initially, model-based designs have been proposed for identifying ODs by incorporating both efficacy and toxicity response. Some frequentist approach has been proposed, for example [2] and [3]. Yuan et al. [15] have provided a comprehensive review of phase I-II trials, particularly Bayesian model-based designs. Recently, model-assisted designs based on efficacy and toxicity have been proposed to identify the ODs in early-phase dose-finding trials. Lin and Yin [16] have proposed the STEIN design based on optimized intervals for toxicity and efficacy. Li et al. [17] have proposed a toxicity and efficacy probability interval (TEPI) design that separately models the toxicity and efficacy to assign new patients to a dose level with a favorable efficacy rate and a tolerable toxicity profile. Takeda et al. [18, 19] have proposed the Bayesian optimal interval design for dose-finding based on both efficacy and toxicity outcomes (BOIN-ET design). Lin et al. [20] have proposed the BOIN12 design, employing a utility function to measure a dose risk–benefit trade-off.

In addition to incorporating both efficacy and toxicity, the late onset of toxicity outcomes has also been a discussion point in many early oncology trials. In immune checkpoint inhibitors, it is well-known that some immune-related toxicity can have a delayed onset; for instance, endocrinopathies occur late and have been observed between weeks 12 and 24 [21]. Notice that late toxicity onset usually requires a larger DLT window to make dose escalation decisions and often causes longer enrollment suspensions. Several designs have been proposed to allow sequential enrollment (i.e., no enrollment suspensions) even if some enrolled patients are still pending DLT assessment to minimize some of the unnecessary long trial recruiting pauses. For dose-finding studies based solely on toxicity, time-to-event CRM (TITE-CRM) [22], rolling six design (R6) [23], time-to-event BOIN design (TITE-BOIN) [24], time-to-event keyboard design (TITE-keyboard) [25], rolling TPI design (R-TPI) [26], and probability-of-decision TPI design (POD-TPI) [27], have been proposed. For the expansion (phase II) part to assess efficacy, Lin et al. [28] have extended the BOP2 design to the time-to-event Bayesian optimal phase II trial (TOP) design by incorporating the pending efficacy data into the posterior distribution of the efficacy outcome. For phase I/II dose-finding, Jin et al. [29] have proposed a general approach of using data augmentation to impute delayed outcomes based on the efficacy–toxicity trade-off design proposed by Thall and Cook [1]. Takeda et al. [30] have extended the BOIN-ET design to the TITE-BOIN-ET design, while Zhou et al.

[31] have extended the BOIN12 design to the TITE-BOIN12 design.

In this paper, we intend to compare the two commonly-used designs for early-phase oncology trials in searching for the final optimal dose. 1) Type A design: conducting only the dose-finding part to identify the OD based on both the efficacy and toxicity outcomes and 2) Type B design: firstly conducting the dose-finding based solely on toxicity, followed by the expansion part to evaluate the efficacy further before a final dose can be determined as the OD. Here the expansion cohorts can be run parallel to the dosing finding cohorts. We focus on the model-assisted approaches because they are more popular due to their performance and simplicity in the actual oncology trials than the model-based designs. We present a simulation study of these approaches to explore the properties across various realistic settings.

The remainder of this article proceeds as follows. In “[Review of Methods](#)” section, the selected designs are briefly reviewed. The simulation results are summarized in “[Numerical Simulations](#)” section. We close with some discussion in “[Discussion](#)” section.

Review of Methods

This section briefly reviews the methods being compared in this paper.

BOIN Design and TITE-BOIN Design

The original BOIN design has gained popularity since it was first proposed by Liu and Yuan [10].

Let $p_1 \leq \dots \leq p_J$ be the true toxicity probability of a set of J dose. Let φ denote the target toxicity probability by the investigator. let y_j and n_j denote the number of patients who experienced toxicities and the number of patients treated at dose level j , respectively. Let $\hat{p}_j = y_j/n_j$ denote the estimated toxicity probability based on all the cumulative data on level j . Liu and Yuan [10] consider the following dose allocation rule. If $\hat{p}_j \leq \lambda_1$, the dose level is escalated to $j+1$; if $\lambda_1 < \hat{p}_j < \lambda_2$, the next cohort continues to be treated at the same dose level j ; if $\lambda_2 \leq \hat{p}_j$, the dose level is de-escalated to $j-1$. The optimal lower and upper boundaries $0 < \lambda_1 \leq \varphi \leq \lambda_2 < 1$ that minimize the posterior probability of incorrect decisions are obtained by

$$\lambda_1 = \log \frac{(1 - \phi_1)}{(1 - \phi)} / \log \frac{\phi(1 - \phi_1)}{\phi_1(1 - \phi)}$$

$$\lambda_2 = \log \frac{(1 - \phi)}{(1 - \phi_2)} / \log \frac{\phi_2(1 - \phi)}{\phi(1 - \phi_2)}$$

Liu and Yuan [10] recommended using $\varphi_1 = 0.6\varphi$ and $\varphi_2 = 1.4\varphi$ as default values. For example, if the target toxicity probability is $\varphi = 0.30$ and the design parameters are $\varphi_1 = 0.6\varphi$ and $\varphi_2 = 1.4\varphi$, the optimal values of (λ_1, λ_2) are derived to be $\lambda_1 = 0.236$ and $\lambda_2 = 0.359$, respectively. The BOIN design is straightforward to implement in actual clinical trials, similar to the 3 + 3 design in practice. During the trial, clinicians count the number of patients who experience toxicity and compare the observed toxicity rate \hat{p}_j with the prespecified interval boundaries λ_1 and λ_2 to determine the next dose assignments until the end of the dose-finding. Subsequently, isotonic regression is applied to estimate the observed toxicity probability at dose j using the pooled adjacent violator algorithm (PAVA) [32] to determine the MTD.

Yuan et al. [24] extended the BOIN design to the time-to-event BOIN (TITE-BOIN) design to minimize enrollment interruption for trial acceleration. The TITE-BOIN design imputes the pending toxicity data using the observed data from all patients as well as the existing data from the pending patients to calculate the toxicity probability at dose level j ,

$$\hat{p}_j = \frac{y_j + \frac{p_j}{1-p_j}(c - STFT)}{n_j}$$

where $STFT = \frac{\sum_{i=1}^c t_i}{W}$ is the standardized total follow-up time, c is the number of pending patents. The unknown parameter p_j will be replaced with its Bayesian posterior mean based on the observed data. For details please see reference paper [24]. As a result, the TITE-BOIN design can use the same optimal boundaries as the BOIN design to make dose assignment decisions for new patients, while some enrolled patients' toxicity data are still pending.

BOP2 and TOP Design

Zhou et al. [14] have proposed a flexible Bayesian optimal phase 2 (BOP2) design to evaluate the preliminary efficacy data for phase 2 oncology trials. The design can handle both simple (e.g., binary) and complicated (e.g., ordinal, nested, and co-primary) endpoints under a unified framework with multiple interim analyses allowed to make flexible trial decisions. The BOP2 design uses the posterior probability of the linear combination of model parameters to make go/no-go decisions. Specifically, let n and N denote the sample size at the interim analysis and the final analysis. Let p_{ORR} denote the overall response rate (ORR) and p_0 denote the inefficacious ORR. The trial is terminated for futility if

$$Pr(p_{ORR} > p_0 | data) < \lambda \left(\frac{n}{N}\right)^\gamma$$

where λ and γ are design tuning parameters to achieve the maximum statistical power while maintaining the desirable

type I error rate. For example, for the design with one interim analysis at 10 and the maximum sample size is 30, assuming that the efficacious ORR is 60% and the inefficacious ORR is 30%, the design parameters are $\lambda = 0.87$ and $\gamma = 0.62$, the optimal stopping boundaries are 2 and 12 at the interim analysis and the final analysis, and the statistical power would be approximately 97% while controlling the type I error rate at 0.1.

Lin et al. [28] extend the BOP2 design to the TOP design to allow real-time “go/no-go” interim decision-making in the presence of late-onset responses by using all available data. The TOP design can utilize pending efficacy information using approximated likelihood and effective sample size (ESS), defined as.

$$ESS = \text{Number of nonpending patients} + \frac{\text{the sum of the follow-up time for pending patients}}{\text{length of the assessment window}}$$

Lin et al. [28] showed that the TOP design could significantly shorten the trial duration.

BOIN-ET Design and TITE-BOIN-ET Design

Takeda et al. [18] have proposed the BOIN-ET design to add the efficacy assessment as part of the dose allocation process. For toxicity, let $p_1 \leq \dots \leq p_J$ be the true toxicity probabilities of a set of J doses for the drug under consideration, and let φ denote the target toxicity probability specified by the investigator. Furthermore, let λ_1 and λ_2 denote the lower and upper cutoffs, satisfying $0 \leq \lambda_1 < \varphi \leq \lambda_2 < 1$. Suppose the current cohort is treated at dose level j , and let y_j and n_j denote the number of patients who experienced toxicities and the number of patients treated at dose level j , respectively. Let $\hat{p}_j = y_j/n_j$ denote the estimated toxicity probability based on all the cumulative data on level j . For efficacy, let q_1, \dots, q_J be the true efficacy probabilities of a set of J doses, and let δ denote the target efficacy probability specified by the investigator. Let η_1 denote the cutoff on the efficacy, satisfying $0 \leq \eta_1 < \delta < 1$. Let x_j denote the number of efficacies at dose level j . $\hat{q}_j = x_j/n_j$ is the estimated efficacy probability at level j . Here we refer Fig. 1 used in the original paper by Takeda et al. [14], which concisely illustrates the dose allocation rule. For the lower middle cell, escalation if the next dose level has not been tried; if the next dose level has been tried, pick the dose level with the highest estimated efficacy probability among the dose level $j - 1, j, j + 1$. The toxicity boundaries λ_1 and λ_2 and efficacy cutoff η_1 is obtained by minimizing the posterior probability of incorrect decisions based on $\varphi, \varphi_1, \varphi_2, \delta$, and δ_1 . Takeda et al. [18] recommend using $\varphi_1 = 0.1\varphi, \varphi_2 = 1.4\varphi$, and $\delta_1 = 0.6\delta$ as default values. The optimal values of $(\lambda_1, \lambda_2, \eta_1)$ are estimated by a grid search. Based on the cumulative efficacy and

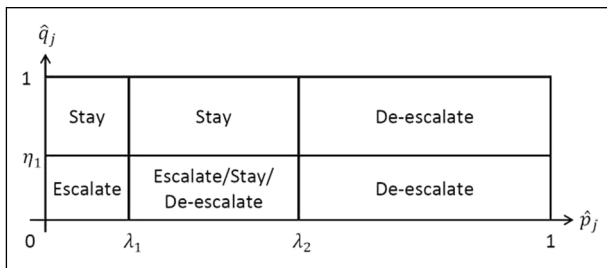


Figure 1 Dose allocation rules for BOIN-ET design considering both efficacy and toxicity.

toxicity outcomes, the BOIN-ET design selects OD at the end of the trial. Takeda et al. [14] used isotonic regression using the PAVA and logistic regression with the fractional polynomial with two degrees of freedom to determine the ODs. Other parametric/non-parametric methods could also be employed in this step.

Takeda et al. [30] have extended the BOIN-ET design to the time-to-event BOIN-ET (TITE-BOIN-ET) design to accelerate the identification of the OD utilizing the likelihood with cumulative and pending data of both efficacy and toxicity outcomes.

BOIN12 Design and TITE-BOIN12 Design

Lin et al. [20] have proposed the BOIN12 design to incorporate efficacy information into the dose selection process. The design uses a utility to measure the toxicity-efficacy trade-off among the possible combination of efficacy and toxicity outcomes and subsequently calculate the desirability at each dose level for these outcomes to define dosing decisions. Specifically, the desirability score (RDS) is estimated from the toxicity-efficacy trade-off utilities based on possible outcomes, i.e., (No Toxicity, Efficacy); (Toxicity, Efficacy); (No Toxicity, No Efficacy); (Toxicity, No Efficacy). Lin et al. [20] show that the RDS can be pre-tabulated before the trial. The algorithm of BOIN12 shares the same escalation and de-escalation boundary as the BOIN design. If $\lambda_2 \leq \hat{p}_j$, the dose level is de-escalated to $j-1$; if $\hat{p}_j \leq \lambda_1$ or $\lambda_1 < \hat{p}_j < \lambda_2$ and $n < 6$, chose the dose level with the highest RDS among $\{j-1, j, j+1\}$ to be the next dose level; if $\lambda_1 < \hat{p}_j < \lambda_2$ and $n \geq 6$, chose the dose level with the highest RDS among $\{j-1, j\}$ to be the next dose level. Based on the cumulative efficacy and toxicity outcomes, the BOIN12 design selects OD at the end of the trial. The MTD is the dose level with the isotonicly estimated toxicity probability closest to the toxicity upper limit. The OD is the dose level that does not exceed the MTD and has the highest RDS.

Zhou et al. [31] have extended the BOIN12 design to the TITE-BOIN12 design considering Bayesian data augmentation and an approximated likelihood method to enable

real-time decision-making when some patients' toxicity and efficacy outcomes are pending.

Numerical Simulations

We carried out simulation studies to evaluate the operating characteristics of designs categorized into two types (A) one-part dose-finding designs based on both efficacy and toxicity outcomes; (B) two parts designs conducting dose-finding part based on toxicity outcome and then opening the expansion part based on efficacy outcomes.

- Type A: (1) BOIN-ET design; (2) TITE-BOIN-ET design; (3) BOIN12 design; (4) TITE-BOIN12 design
- Type B: (5) BOIN and BOP2 designs; (6) TITE-BOIN and BOP2 designs; (7) BOIN and TOP designs; (8) TITE-BOIN and TOP designs

Simulation Settings

As summarized in Fig. 2, to avoid cherry-picking scenarios that favored a specific design, we adopted 12 different scenarios for the true probabilities of efficacy q_j and toxicity p_j at the dose levels used in Sato et al. [33]. In every scenario, the dose–toxicity curve monotonically increased, and the dose–efficacy relationships were monotone, plateau, and bell shape. For all designs, in the dose-finding part, we set the target toxicity at 0.3, the cohort size was three subjects per cohort, and skipping doses was not allowed. The accrual rate was one patient every two weeks, the toxicity observation period was 28 days, and the efficacy observation period was 56 days. The trials were simulated 1,000 times.

Although different designs had different perspectives, we considered the settings to compare the designs as possible as we could fairly. For type A design, i.e., design 1)–4), the maximum sample size was 126 with a stopping rule if the number of patients at the current dose reached 30. For the BOIN-ET and TITE-BOIN designs, the target efficacy was set at 0.6. The optimal values of $(\lambda_1, \lambda_2, \eta_1)$ were $\lambda_1 = 0.14$, $\lambda_2 = 0.35$, and $\eta_1 = 0.48$. The dose level satisfying the tolerability based on isotonic regression using the PAVA and maximizing the efficacy probability based on logistic regression with the fractional polynomial with two degrees of freedom was selected as the OD, as described in “BOIN-ET Design and TITE-BOIN-ET Design” section. For the BOIN12 and TITE-BOIN12 designs, the optimal values of (λ_1, λ_2) were $\lambda_1 = 0.236$ and $\lambda_2 = 0.359$, and the toxicity-efficacy trade-off utilities were set as (No Toxicity, Efficacy) = 100; (Toxicity, Efficacy) = 60; (No Toxicity, No Efficacy) = 40; (Toxicity, No Efficacy) = 0. The dose level that did not exceed the MTD and the highest RDS was selected as the OD, as described in “BOIN12 Design and

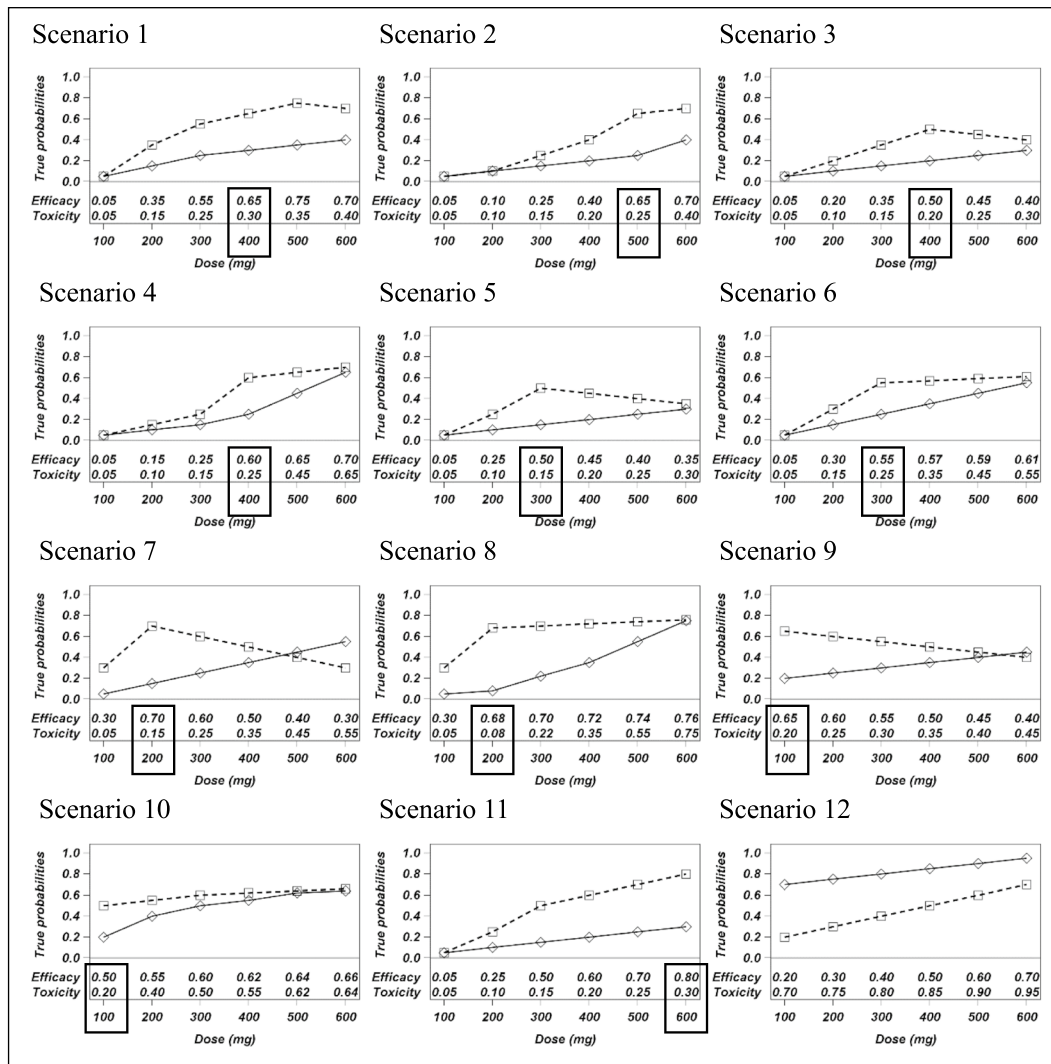


Figure 2 Dose-toxicity and dose-efficacy relationships considered in the simulation study. The true probabilities of efficacy (dotted line) and toxicity (solid line) for each dose level. The optimal doses are in squares.

TITE-BOIN12 Design” section. Stopping criteria were set to avoid allocating patients to ineffective or severely toxic doses. If the posterior probability of efficacy outcome is less than the lower limit of efficacy 0.20 is more than 0.90, the dose levels j are eliminated from the trial. If the posterior probability of toxicity outcome is more than the upper limit of toxicity 0.40 is more than 0.90, the dose levels j and higher are eliminated from the trial. The trial is terminated if all dose levels are eliminated. We assume that the efficacy and toxicity endpoints follow vague beta priors beta(1,1).

For type B designs, i.e., design 5)-8), the dose-finding part using the BOIN and TITE-BOIN designs had eight cohorts with a stopping rule, that is, if nine subjects were assigned to a single dose level and the decision was to stay, the trial would be stopped. The optimal values of (λ_1, λ_2) were $\lambda_1 = 0.236$ and $\lambda_2 = 0.359$, as described in “BOIN

Design and TITE-BOIN Design”. Stopping criteria were set to avoid allocating patients to severely toxic doses. If the posterior probability of toxicity outcome is more than the target toxicity probability 0.30 is more than 0.95, the dose levels j and higher are eliminated from the trial. The trial is terminated if all dose levels are eliminated. We assume that the toxicity endpoints follow vague beta priors beta(1,1). The expansion part used the BOP2 and TOP designs described in “BOP2 and TOP Design”. The sample sizes for the interim and final analysis were 10 subjects and 30 subjects, respectively. Assuming the inefficacious ORR was 0.3 and the efficacious ORR was 0.6, the optimal stopping boundaries were estimated to be 3 and 8 at the interim and final analyses while controlling for 0.1 type I error and achieving 0.97 power. The expansion cohort opened at the dose level in which a response was observed and at

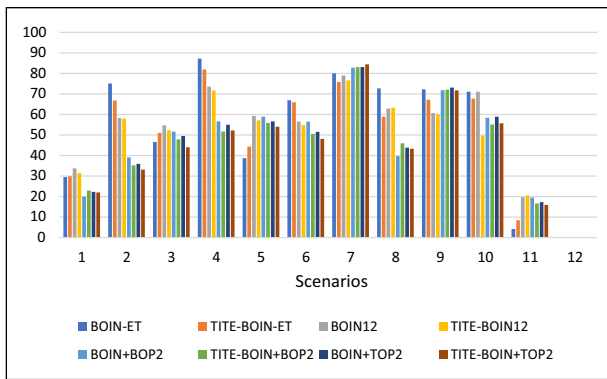


Figure 3 Percentage of times the optimal dose was correctly selected.

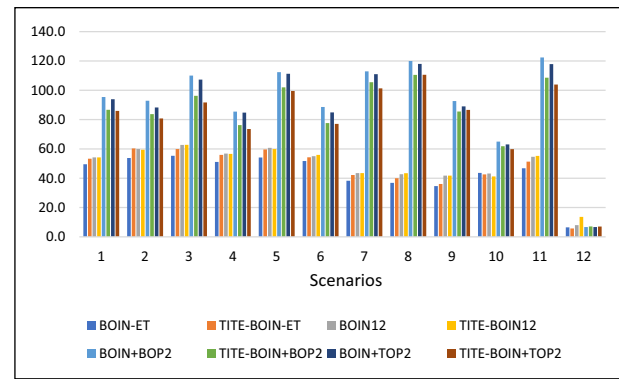


Figure 5 Average number of patients per trial.

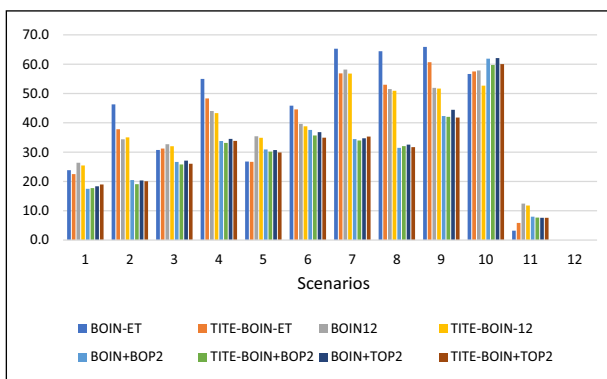


Figure 4 Percentage of patients allocated to the optimal dose.

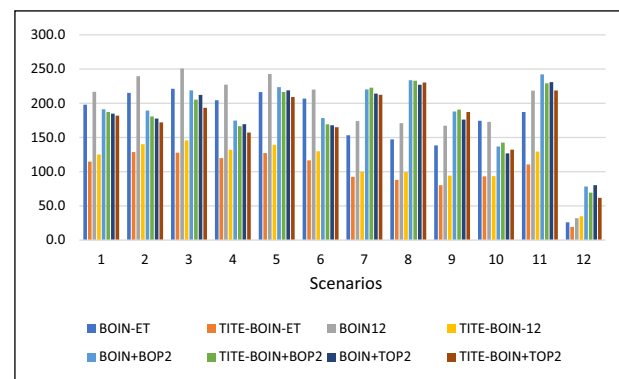


Figure 6 Average trial duration in weeks.

all subsequent dose levels once each dose level had been deemed tolerable. Subjects in the expansion cohort contributed to the tolerability assessment and dose escalation decisions in the escalation part of the trial. The OD was selected by fitting a unimodal isotonic regression using efficacy information from both dose-escalation and expansion cohorts. The unimodal isotonic regression allowed the estimation of the umbrella-shaped dose–response curve. Only those dose levels that had not crossed the optimal stopping boundaries of the BOP2 and TOP designs were considered in the process of OD selection.

Simulation Results

Each simulation under certain scenarios can be viewed as a hypothetical study. The results of the simulation gave us the operating characteristic of each design under a certain scenario and can be used to guide us to select the right design for our study and later to help us select the Optimal Dose for the compound. The operating characteristics of the designs were summarized in Figs. 3, 4, 5, and 6 and Table S1–S4, which were organized by scenario. The results were summarized in terms of the percentage of times the optimal dose was correctly selected, the percentage of patients allocated

to the optimal dose, the average number of patients per trial, and the average trial duration.

Figure 3 and Table S1 showed the percentage of times the optimal dose was correctly selected for the eight designs under comparison. Under Scenarios 1 and 11, the lower doses had sufficient efficacy. In Scenario 1, the type A designs outperformed the type B designs. In Scenario 11, the BOIN12, TITE-BOIN-12, and the type B designs showed similar performance. The BOIN-ET and TITE-BOIN-ET designs selected the correct ODs at lower percentages because the BOIN-ET and TITE-BOIN-ET designs tended to select lower doses as the ODs. Under Scenarios 2 and 4, the dose–efficacy curve monotonically increased. The BOIN-ET and TITE-BOIN-ET designs outperformed the other designs. Compared with the type B designs, the percentage of correct OD selection by the type A designs were higher by more than 15%. Under Scenarios 3, 5, and 7, the dose–efficacy curve non-monotonically increased. In Scenarios 3 and 7, all designs showed similar performance. In Scenario 5, the BOIN-ET and TITE-BOIN-ET designs selected sub-optimal doses as the ODs compared with other designs. Under Scenarios 6 and 8, the dose–efficacy curve reached a plateau. In Scenario 6, the BOIN-ET and TITE-BOIN-ET designs captured the correct ODs at higher percentages than the other designs.

In Scenario 8, the type A designs selected the correct ODs higher by more than 15% than the type B designs. Under Scenarios 9 and 10, the lowest dose level was only tolerable, and the dose-efficacy curve monotonically decreased. In Scenario 9, the BOIN-ET and the type B designs provided similar performance. In Scenario 10, the BOIN-ET, TITE-BOIN-ET, and BOIN12 designs showed higher percentages of correct OD selection. Under scenario 12, all dose levels were too toxic. All designs selected early stopping.

Figure 4 and Table S2 showed the percentage of patients allocated to the optimal dose. In Scenarios 1, 2, 3, 4, 6, 7, 8, and 9, the type A designs, especially the BOIN-ET design, allocated greater percentage of patients to the optimal doses. In Scenarios 5 and 11, the BOIN12 and TITE-BOIN12 designs allocated more patients to the optimal doses.

Figure 5 and Table S3 showed the average number of patients per trial. In All Scenarios except for Scenario 12, the type A designs, especially the BOIN-ET design, needed fewer patients to detect the OD than the type B designs.

Figure 6 and Table S4 showed the average trial duration in weeks. In all Scenarios except for scenario 12, the TITE-BOIN-ET and TITE-BOIN12 designs significantly shortened the average trial duration compared with the type A designs without sequential enrollment and the type B designs. In Scenario 12, the type A designs shortened the average trial duration than the type B designs. Within type B designs, the TITE-BOIN and TOP designs shortened the average trial duration than the other type B designs.

In summary, in most scenarios, the type A designs outperformed the type B designs in correctly selecting the ODs, efficiently assigning patients to the ODs, and reducing the total sample size. In terms of the trial duration, the TITE-BOIN-ET and TITE-BOIN12 designs in type A had the shortest trial duration. Within the type A designs, the BOIN-ET and BOIN12 designs slightly outperformed the TITE-BOIN-ET and TITE-BOIN12 designs in terms of correctly selecting the ODs and efficiently assigning patients to the ODs for many scenarios. On the other hand, the TITE-BOIN-ET and TITE-BOIN12 designs significantly shortened the average trial duration across different scenarios compared with the BOIN-ET and BOIN12 designs. The type B designs were generally less efficient than the type A designs, although a slightly higher percentage of OD selection was observed in Scenario 7 and 9; substantially more subjects and a longer trial duration were needed under Scenarios 7 and 9 for these designs.

Discussion

We compare the operating characteristics of optimal dose-finding approaches in early-phase oncology trials, including multiple expansion cohorts. The designs that utilize

both efficacy and toxicity to guide their dose-finding as a one-part design has shown superior performance in term of the higher percentage of correct selection of OD, the higher percentage of allocating patients to the OD, and reducing the total sample size in most cases. However, the BOIN-ET and BOIN12 designs may result in long trial duration in some cases, as shown in the figure on trial duration. The TITE version of these two designs can significantly reduce the trial duration. On the other hand, the TITE assumption does not work well if efficacy readouts are delayed. So a more real-time readout of the efficacy outcome, once it occurs, is required in practice. We also want to point out that with the dose-finding part, followed by the dose-expansion part approach, e.g., the combination of the BOIN and the BOP2 designs, additional information is gained from the BOP2 design interim and final analysis. Therefore, there is a lower chance of selecting a sub-therapeutic dose level as the OD because the OD selection will be based on those dose levels considered promising at the end of the BOP2 or TOP designs. In addition, the type B designs are still valuable if a dose-finding trial with expansion cohorts wants to strictly control the type I error rate for efficacy evaluation. Our simulation result has some limitations due to limited settings and scenarios. Although the current setting of 28 days and 56 days of safety and efficacy assessment windows is one of the practical settings, depending on the compound and indication, the safety and efficacy assessment window might be different, and these can possibly change the operating characteristic of the designs. For example, if the window of the outcome assessment for efficacy is much longer than 56 days, the average trial duration for the designs in type A might be longer, and the accuracy of the estimation of OD might be worse in the TITE approaches. In addition, in type A designs, the BOIN-ET design and the BOIN12 design have different OD selection criteria. The difference may affect the performance of correct OD selection. Therefore, it is important to conduct a comprehensive simulation before the trial to understand the operating characteristics. We would like to advocate type A design based on the simulation result we see in this paper and strongly encourage statisticians to include type A designs and compare type A designs and type B designs in their trial design operating characteristic simulations. As type A design is relatively new, real clinical trial examples utilizing type A design is limited. Therefore, it is valuable to publish the results and findings when the type A designs are implemented in the actual clinical trials.

In recent years, many new methods have been proposed for the oncology early phase dose-finding trial. Statisticians should work with clinicians to carefully consider candidate designs. Furthermore, willing to take an innovative approach

to facilitate better and accelerate trials. Case-by-case simulation of the novel designs is essential in this process.

Conclusion

The operating characteristics show that the designs that utilized toxicity and efficacy simultaneously are superior in selecting the ODs, efficiently assigning patients to the ODs, and reducing the total sample size to designs that utilized toxicity and efficacy sequentially under most scenarios. In addition, the TITE version of designs that utilized toxicity and efficacy simultaneously with sequential enrollment significantly reduce the trial duration compared with other designs.

Author Contributions

RL: Substantial Contribution to the concept, design, conduction of the work. Drafting the work and revising it. KT: Substantial Contribution to the design, conduction of the work. Revising the draft Critically for important intellectual content AR: Revising the draft critically for important intellectual contents.

Funding

There are no financial holdings or considerations that could influence the objective of the report.

Data Availability

R shiny app available on <https://www.trialdesign.org/> for BOIN12 and TITE-BOIN12 design; SAS code for BOIN-ET and TITE-BOIN-ET and R code for BOIN + BOP2 and TITE-BOIN + BOP2 available upon request.

Declarations

Conflict of interest

The authors have nothing to disclose.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s43441-023-00517-1>.

References

1. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*. 2004;60(3):684–93.
2. Ivanova A. A new dose-finding design for bivariate outcomes. *Biometrics*. 2003;59(4):1001–7.
3. Dragalin V, Fedorov V. Adaptive designs for dose-finding based on efficacy–toxicity response. *J Stat Plan Inference*. 2006;136(6):1800–23.
4. FDA (2018) Expansion cohorts: Use in first-in-human clinical trials to expedite development of oncology drugs and biologics guidance for industry. URL:<https://www.fda.gov/media/115172/download>.
5. O’Quigley J. Sequential monitoring of Phase I dose expansion cohorts. *Stat Med*. 2017;36(2):204–14.
6. Storer BE. Design and analysis of phase I clinical trials. *Biometrics*. 1989;45(3):925–37.
7. O’Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics*. 1990;46(1):33–48.
8. Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Stat Med*. 2008;27(13):2420–39.
9. Ji Y, Wang SJ. Modified toxicity probability interval design: a safer and more reliable method than the 3 + 3 design for practical phase I trials. *J Clin Oncol*. 2013;31(14):1785–91.
10. Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *J R Stat Soc Ser C*. 2015;64(3):507–23.
11. Yan F, Mandrekar SJ, Yuan Y. Keyboard: a novel bayesian toxicity probability interval design for phase I clinical trials. *Clin Cancer Res : An Off J Am Assoc Cancer Res*. 2017;23(15):3994–4003.
12. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989;10(1):1–10.
13. Thall PF, Simon R. A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials. *Control Clin Trials*. 1994;15(6):463–81.
14. Zhou H, Lee JJ, Yuan Y. BOP2: Bayesian optimal design for phase II clinical trials with simple and complex endpoints. *Stat Med*. 2017;36(21):3302–14.
15. Yuan Y, Nguyen HQ, and Thall PF. Bayesian designs for phase I–II clinical trials. Chapman & Hall/CRC Biostatistics Series. CRC Press; 2016.
16. Lin R, Yin G. STEIN: a simple toxicity and efficacy interval design for seamless phase I/II clinical trials. *Stat Med*. 2017;36(26):4106–20.
17. Li DH, Whitmore JB, Guo W, Ji Y. Toxicity and efficacy probability interval design for phase I adoptive cell therapy dose-finding clinical trials. *Clin Cancer Res*. 2017;23(1):13–20.
18. 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(4):383–95.
19. Takeda K, Morita S, Taguri M. gBOIN-ET: the generalized Bayesian optimal interval design for optimal dose-finding accounting for ordinal graded efficacy and toxicity in early clinical trials. *Biom J*. 2022;64(7):1178–91.
20. Lin R, Zhou Y, Yan F, Li D, Yuan Y. BOIN12: Bayesian optimal interval Phase I/II trial design for utility-based dose finding with immunotherapy and targeted therapies. *JCO Precis Oncol*. 2020;4:1393–402.
21. Weber JS, D’Angelo SP, Minor D, Hodi FS, Gutzmer R, Neyns B, Hoeller C, Khushalani NI, Miller WH Jr, Lao CD, Linette GP, Thomas L, Lorigan P, Grossmann KF, Hassel JC, Maio M, Sznol M, Ascierto PA, Mohr P, Chmielowski B, Bryce A, Svane IM, Grob JJ, Krackhardt AM, Horak C, Lambert A, Yang AS, Larkin J. Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol*. 2015;16(4):375–84.
22. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*. 2000;56(4):1177–82.
23. Skolnik JM, Barrett JS, Jayaraman B, Patel D, Adamson PC. Shortening the timeline of pediatric phase I trials: the

- rolling six design. *J Clin Oncol : Off J Am Soc Clin Oncol*. 2008;26(2):190–5.
24. Yuan Y, Lin R, Li D, Nie L, Warren KE. Time-to-event Bayesian optimal interval design to accelerate phase I trials. *Clin Cancer Res*. 2018;24(20):4921–30.
 25. Lin R, Yuan Y. Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. *Biostatistics*. 2020;21(4):807–24.
 26. Guo W, Ji Y, Li D. R-TPI: rolling toxicity probability interval design to shorten the duration and maintain safety of phase I trials. *J Biopharm Stat*. 2019;29(3):411–24.
 27. Zhou T, Guo W, Ji Y. PoD-TPI: probability-of-decision toxicity probability interval design to accelerate phase I trials. *Stat Biosci*. 2020;12:124–45.
 28. Lin Ruitao, Coleman Robert L, Yuan Ying. TOP: time-to-Event bayesian optimal phase II trial design for cancer immunotherapy. *JNCI: J Natl Cancer Inst*. 2020;112(1):38–45.
 29. Jin IH, Liu S, Thall PF, Yuan Y. Using data augmentation to facilitate conduct of phase I-II clinical trials with delayed outcomes. *J Am Stat Assoc*. 2014;109(506):525–36.
 30. Takeda K, Morita S, Taguri M. TITE-BOIN-ET: Time-to-event Bayesian optimal interval design to accelerate dose-finding based on both efficacy and toxicity outcomes. *Pharm Stat*. 2020;19(3):335–49.
 31. Zhou Y, Lin R, Lee JJ, et al. TITE-BOIN12: a Bayesian phase I/II trial design to find the optimal biological dose with late-onset toxicity and efficacy. *Stat Med*. 2022;41(11):1918–31.
 32. Barlow RE, Bartholomew D, Bremner JM, Brunk HD. *Statistical inference under order restrictions; the theory and application of isotonic regression*. New York: Wiley; 1972.
 33. Sato H, Hirakawa A, Hamada C. An adaptive dose-finding method using a change-point model for molecularly targeted agents in phase I trials. *Stat Med*. 2016;31(6):516–32.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.