Modified isotonic regression based phase I/II clinical trial design identifying optimal biological dose

Yingjie Qiu[1], Yi Zhao[1], Hao Liu[2], Sha Cao[1,3], Chi Zhang[3,4] and Yong Zang[1*,3]

1. Department of Biostatistics and Health Data Science, Indiana University, USA

2. Department of Biostatistics and Epidemiology, Cancer Institute of New Jersey, Rutgers University, USA

3. Center of Computational Biology and Bioinformatics, Indiana University, USA

4. Department of Medical and Molecular Genetics, Indiana University, USA

*For correspondence: Yong Zang, Department of Biostatistics and Health Data Science, Indiana University, 410 W 10th Street, Indianapolis, IN, 46202, US. Email: zangy@iu.edu

Abstract

Conventional phase I/II clinical trial designs often use complicated parametric models to characterize the dose-response relationships and conduct the trials. However, the parametric models are hard to justify in practice, and the misspecification of parametric models can lead to substantially undesirable performances in phase I/II trials. Moreover, it is difficult for the physicians conducting phase I/II trials to clinically interpret the parameters of these complicated models, and such significant learning costs impede the translation of novel statistical designs into practical trial implementation. To solve these issues, we propose a transparent and efficient phase I/II clinical trial design, referred to as the modified isotonic regression-based design (mISO), to identify the optimal biological doses for molecularly targeted agents and immunotherapy. The mISO design makes no parametric model assumptions on the dose-response relationship and yields desirable performances under any clinically meaningful dose-response curves. The concise, clinically interpretable dose-response

models and dose-finding algorithm make the proposed designs highly translational from the statistical community to the clinical community. We further extend the mISO design and develop the mISO-B design to handle the delayed outcomes. Our comprehensive simulation studies show that the mISO and mISO-B designs are highly efficient in optimal biological dose selection and patients allocation and outperform many existing phase I/II clinical trial designs. We also provide a trial example to illustrate the practical implementation of the proposed designs. The software for simulation and trial implementation are available for free download.

Keywords: Phase I/II clinical trials; Molecularly targeted agents; Immunotherapy; Delayed outcome; Isotonic regression.

1 Introduction

Molecularly targeted agents (MTA) and immunotherapy (IT) have revolutionized cancer treatment and represent the most promising new treatment in recent decades for almost any kind of cancer. MTA is developed to modulate specific aberrant pathways in cancer cells while sparing normal tissue [1, 2]. IT stimulate a patient's immune system to fight cancer [3, 4, 5, 6]. The monotonically increasing dose-efficacy curve assumption has been challenged by the emergence of MTA/IT [7, 8, 9, 10, 11]. As a result, traditional phase I dose-finding designs identifying the maximum tolerated dose (MTD) are not suitable for MTA/IT because the optimal therapeutic effects are not necessarily observed at the MTD. Therefore, new dose-finding designs are required for clinical trials evaluating MTA/IT, which target the optimal biological dose (OBD), maximizing the overall therapeutic effects considering both toxicity and efficacy.

There is a rich body of literature on phase I/II clinical trial designs existing in the statistical community and most are model-based designs [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. Various parametric models have been used to characterize the dose-toxicity and dose-efficacy curves (logistic regression, change-point model, copula model, etc.), with different model assumptions. It is known that the performances of model-based dose-finding designs are highly sensitive to the underlying parametric model assumptions [29], which are difficult to verify in practice. Hence, physicians need to justify the underlying model assumptions and comprehend the statistical models before selecting a suitable design. Also, the computational burden and heavy involvement of statistical consultation cause additional administrative encumbrance. These high learning costs have substantially impeded the implementation of statistically innovative phase I/II designs in clinical practice. We have reviewed 9 phase I/II MTA/IT trials published in the journal *Clinical Cancer Research* from January 2018 to September 2019 [30, 31, 32, 33, 34, 35, 36, 37, 38]. All the trials used the conventional paradigm of finding the MTD first and then conducting cohort expansion. None of the trials used both toxicity and efficacy jointly to decide patients' allocation and dose selection.

This paper develops a transparent yet efficient phase I/II clinical trial design, referred to as the modified isotonic regression-based design (mISO), to identify the OBD for MTA/IT by simultaneously monitoring efficacy and toxicity outcomes in a single trial. For most MTA/IT, The monotonically increasing dose-efficacy curve assumption does not hold. Instead, the efficacy response often increases initially with the dose and then plateaus [26, 28]. Indeed, from the pharmacological perspective, MTA works on the biomarker pathway, and IT works on the immune system. So once the biomarker pathway has been saturated for MTA or the immune system has been fully boosted, further increasing dose levels generally does not yield a better response. For example, nivolumab is a programmed death–1 inhibitor used to treat melanoma, renal cell carcinoma (RCC), and non-small cell lung cancer

(NSCLC) as MTA. A comprehensive nivolumab dose selection study revealed that an initially increasing dose-objective tumor response trend was observed for each tumor type but eventually appeared to plateau at nivolumab doses of ≥ 1 mg/kg for melanoma and RCC and at ≥ 3 mg/kg for NSCLC [39]. The mISO design is based on the plateaued dose-response curve supported by pharmacological mechanisms and historical studies. However, this design makes no parametric model assumption. The idea is to extend the pooled–adjacent-violators algorithm (PAVA) [40] such that a modified version of the non-parametric isotonic regression can be used to fit the plateaued curve and identify the change point. Without the complex parametric model assumptions, the mISO design is highly transparent and can be easily understood and used by the clinical community.

Delayed outcomes are common in the clinical study where patients who have not reported responses at the moment of dose allocation may report the responses later during the remaining follow-up. In phase I trials, the delayed outcome can result in underestimating the toxicity probabilities and allocating an undesirably large number of patients at overly toxic doses. This problem becomes more intractable in phase I/II clinical trials because either toxicity or efficacy outcomes may be subject to delayed response. We also investigate this issue and extend the mISO design to handle the delayed outcomes.

Our research is inspired by an emerging phase I/II clinical trials at the Indiana University Melvin and Bren Simon Comprehensive Cancer Center (IUSCCC). The goal is to determine the optimal dose of a TIM-3 inhibitor for the adolescent and young adult (AYA) population of 1-39 year olds with relapsed acute myelogenous leukemia (AML). Toxicity will be graded according to the NCI-CTCAE version 4.03, and tumor response will be evaluated using RECIST version 1.1. As an MTA, the objective tumor response rate of the TIM-3 inhibitor does not monotonically increase with the dose. Instead, the plateaued dose-response pattern has been repeatedly identified in similar trials previously conducted in IUSCCC. Also, the assessment window is 3 months for both the toxicity and efficacy outcomes, whereas on

average, 3 patients will enter the trial every 1 month, causing the delayed outcome issue. Therefore, how to design a phase I/II trial with plateaued dose-response curve and delayed outcomes is the challenge of this study, which motivates us to develop the mISO design.

We have previously developed an isotonic regression-based design (ISO) for phase I/II clinical trials [18]. The differences between the previous ISO design and the new mISO design developed in this paper are listed below. (1) The ISO design relies on the umbrella shape dose-response curve, and the mISO is based on a clinically more relevant plateaued dose-response curve. (2) The ISO design is for immediately observable outcomes only, and the mISO design can handle both the immediately observable and delayed outcomes. (3) The ISO has no early stopping rules, and the mISO can early terminate the trial due to either overly toxic or inefficacious doses. (4) the ISO design uses a "sum of squared error" criterion to determine the location of OBD. The mISO design uses the more efficient AIC criterion (supported by the sensitivity analysis).

2 Probability model

Let $D = (d_1, \cdots, d_J)$ denote a set of J prespecified increasing doses under investigation. Let $T = 0, 1$ be the binary dose-limiting toxicity (DLT) endpoint and $E = 0, 1$ be the binary efficacy endpoint (e.g., objective tumor response). We define $p_{T,j} = \Pr(T = 1|D = d_j)$ as the DLT rate and $p_{E,j} = \Pr(E = 1|D = d_j)$ as the efficacy rate, both at dose d_j for $j = 1, \cdots, J$. The toxicity rates generally monotonically increase with the doses so

$$p_{T,1} < \cdots < p_{T,J}. \tag{1}$$

Based on the rationale provided in the Introduction section, we specify a plateaued dose-

efficacy curve such that

$$p_{E,1} < \cdots < p_{E,j^\dagger} = p_{E,j^\dagger+1} = \cdots p_{E,J}. \tag{2}$$

Let ϕ_T be the highest acceptable DLT rate and ϕ_E be the lowest acceptable efficacy rate. To exclude overly toxic and less efficacious doses, we define the admissible set of toxicity $\mathcal{A}_T$ including the doses with acceptable DLT rates (no greater than ϕ_T) and the admissible set of efficacy $\mathcal{A}_E$ including the doses with acceptable efficacy rates (no less than ϕ_E). Then, we restrict the OBD within the overall admissible set $\mathcal{A} = \mathcal{A}_T \cap \mathcal{A}_E$. Finally, we define OBD as the dose that yields the highest efficacy rate within the admissible set. If multiple doses satisfy this definition, the lowest one is selected as OBD. This definition has also been used by many other phase I/II clinical trial designs [18, 26, 41, 42], mainly due to its clinically meaning interpretation.

Let n_j be the number of patients treated at dose j, and $r_{T,j}$ and $r_{E,j}$ be the number of patients experiencing $T = 1$ and $E = 1$ among the n_j patients, respectively. To construct the admissible set, let $\mathcal{O}$ be the observed data, at each dose d_j, we propose the following beta-binomial model for each dose D_j independently:

$$\begin{aligned}
&p_{T,j} \sim \text{Beta}(\alpha_T, \beta_T); \quad p_{E,j} \sim \text{Beta}(\alpha_E, \beta_E), \\
&r_{T,j}|p_{T,j} \sim \text{Bin}(n_j, p_{T,j}); \quad r_{E,j}|p_{E,j} \sim \text{Bin}(n_j, p_{E,j}), \\
&p_{T,j}|\mathcal{O} \sim \text{Beta}(\alpha_T + r_{T,j}, \beta_T + n_j - r_{T,j}); \quad p_{E,j}|\mathcal{O} \sim \text{Beta}(\alpha_E + r_{E,j}, \beta_E + n_j - r_{E,j}).
\end{aligned} \tag{3}$$

Then, let μ_T be the pre-determined cut-off probability value for toxicity (e.g., 0.9 or 0.95), we claim that a dose D_j is overly toxic if $\Pr(p_{T,j} > \phi_T|\mathcal{O}) > \mu_T$, i.e., the chance that

the DLT rate $p_{T,j}$ is greater than the maximum acceptable value ϕ_T is high. Based on (1), if a dose D_j is overly toxic, then all the doses beyond D_j are overly toxic. Therefore, let D_{j_T} be the lowest dose that is overly toxic, i.e., $D_{j_T} = \text{argmin}_{D_j}\big(\text{pr}(p_{T,j} > \phi_T|\mathcal{O}) > \mu_T\big)$. The admissible set of toxicity $\mathcal{A}_T$ is constructed as $\mathcal{A}_T = \{D_1 : D_{j_T-1}\}$. Along the same line, let μ_E be the pre-determined cut-off probability value for efficacy. Based on (2), we find $D_{j_E} = \text{argmax}_{D_j}\big(\text{pr}(p_{E,j} < \phi_E|\mathcal{O}) > \mu_E\big)$ and construct the admissible set of efficacy as $\mathcal{A}_E = \{D_{j_E+1} : D_J\}$.

After constructing the admissible set, the next step is to estimate $p_{E,j}$. First of all, if we only have the left monotonically increasing part of pattern (2) as $p_{E,1} < \cdots < p_{E,j^\dagger}$, then the conventional isotonic regression method can be used, which applies the pooled–adjacent-violators algorithm (PAVA) [40] to estimate $p_{E,j}$. In particular, let $\hat{p}_{E,j} = r_{E,j}/n_j$ be the proportional estimate for $p_{E,j}$, the PAVA replaces any adjacent $\hat{p}_{E,j}$'s that violate the non-decreasing order by their (weighted) average so that the resulting isotonic estimates $\tilde{p}_{E,j}$ become monotonic.

The PAVA is fast and transparent, and the resulting isotonic estimates are the maximum likelihood estimates (MLE) under the monotonically increasing order constraint [43]. However, a direct application of the PAVA to our research topic is inefficient because the unique right part of the pattern (2) $p_{E,j^\dagger} = p_{E,j^\dagger+1} = \cdots p_{E,J}$. We propose a "collapse-then-split" method to solve this problem. We first assume that the location of $j^\dagger$ is known and will relax this assumption later. When $j^\dagger$ is known, We can artificially define a "super" dose $j^\triangle$ combing all the doses in the plateau $(D_{j^\dagger}, \cdots, D_J)$. Then, we (1) derive the proportional estimates as $\hat{p}_{E,1} = r_{E,1}/n_1, \cdots, \hat{p}_{E,j^\triangle} = \sum_{j=j^\dagger}^{J} r_{E,j} / \sum_{j=j^\dagger}^{J} n_j$, (2) apply the PAVA to $\{\hat{p}_{E,1}, \cdots, \hat{p}_{E,j^\triangle}\}$ and get the isotonic estimates $\{\tilde{p}_{E,1}, \cdots, \tilde{p}_{E,j^\triangle}\}$, and (3) re-define the isotonic estimates in plateau as $\tilde{p}_{E,j^\dagger} = \cdots = \tilde{p}_{E,J} = \tilde{p}_{E,j^\triangle}$.

The "collapse-then-split" algorithm cannot be directly used when the location of $j^\dagger$ is unknown. We extend the original isotonic regression method and propose the modified

isotonic regression method to circumvent this issue. The key is that, although we cannot pre-specify the location of $j^\dagger$, the choices are limited because there are only a few doses under investigation, and one of them must be $j^\dagger$. We can enumerate all the possible locations for $j^\dagger$. For each specification of $j^\dagger$, we can derive the corresponding isotonic estimates. Then, we can select the best location and the corresponding estimates based on a statistical model-fitting criterion (e.g., AIC). In particular, the modified isotonic regression method is summarized as follows.

1. Enumerate all the possible locations of $j^\dagger$ listed as $j^\dagger = 1, \cdots, J$.

2. For every $j^\dagger = l$ with $l = 1, \cdots, J$, using the proposed "collapse-then-split" algorithm to get the corresponding isotonic estimates $\{\tilde{p}_{E,j}^{(l)}, j = 1, \cdots, J\}$ and the associated AIC referred to as $\text{AIC}^{(l)}$.

3. Select the best location of $j^\dagger$ as $j^\dagger = l^\star$ yielding the smallest AIC value, i.e., $l^\star = \text{argmin}_{l \in (1, \cdots, J)} \text{AIC}^{(l)}$.

4. Select the corresponding set of estimates $\{\tilde{p}_{E,j}^{(l^\star)}, j = 1, \cdots, J\}$ as the final modified isotonic estimates.

3 Dose-finding design

We can now develop the dose-finding design with the admissible set and modified isotonic estimates at hand. We note that as a non-parametric method, we can only estimate the tried doses. That is, let D_{j° be the highest dose that has been tried, the proposed admissible set and modified isotonic regression estimates can only be applied to the doses $\{D_1, \cdots, D_{j^\circ}\}$. Therefore, if we restrict our dose-finding procedure within the tired doses, we may identify a local OBD and miss the global optimal value. For example, we consider a hypothetical

trial by treating the first cohort of patients at the lowest dose D_1, which yields acceptable DLT and efficacy rates. Then, because $j^\circ = 1$, the admissible set contains one dose D_1, so D_1 is the identified OBD, and the next cohort of patients will still be treated at D_1. The procedure repeats, and the dose-finding will be restricted to D_1 through the trial until the maximum sample size is reached. To break this deadlock and identify the global OBD, we propose that if the highest tried dose D_{j° is safe and j° is not the highest dose, we will escalate the dose to explore more untried doses.

The proposed dose-finding design can be summarized as follows:

1. We treat the first cohort of patients at the lowest dose D_1, or at the physician-specified dose.

2. We identify the highest tried dose D_{j°. If $\Pr(p_{T,j^\circ} > \phi_T|\mathcal{O}) \leq \mu_T$ and $j^\circ < J$, we treat the next cohort of patients at dose $D_{j^\circ+1}$ and go to step 5; otherwise, we go to step 3.

3. We construct the admissible set $\mathcal{A}$ using the proposed beta-binomial model. If $\mathcal{A}$ is empty, we terminate the trial and claim that no dose should be selected; otherwise, we go to step 4.

4. We use the proposed modified isotonic regression method to estimate the efficacy rates for all the doses within $\mathcal{A}$. The dose level yielding the highest efficacy rate estimate is the identified OBD level. Let j be the current dose level. If OBD $< j$, we treat the next cohort of patients at dose D_{j-1}; if OBD $= j$, we treat the next cohort of patients at dose D_j; if OBD $> j$, we treat the next cohort of patients at dose D_{j+1}.

5. We repeat steps 2-4 until the maximum sample size has been reached.

At the end of the trial, we construct the final admissible set and calculate the efficacy rates for all the doses within the admissible set using all the data. The dose that yields the

highest efficacy rate estimate within the admissible set is recommended as the final OBD. Because the proposed design relies on the modified isotonic regression, we refer to it as the modified isotonic regression-based dose-finding design (mISO) hereafter.

4 Delayed outcomes

The mISO design requires that both the toxicity and efficacy outcomes can be ascertained quickly after the initiation of the treatment to inform the dose assignment for the next cohort of patients entering the trial. In other words, the toxicity and efficacy outcomes should be completely observable for all the patients treated in the trial by the time of interim analysis for new dose assignment. However, delayed outcomes are common in dose-finding trials. In many circumstances, it is unlikely to have the response outcomes for all the previous patients being observable before another cohort of patients are enrolled in the trial. Indeed, patients who have not experienced toxicity or efficacy events during decision-making may still experience such events later within the assessment window.

In a phase I/II trial, patients enter the trial sequentially and are followed for a fixed period $[0, U_T]$ for binary toxicity outcome T and a fixed period $[0, U_E]$ for binary efficacy outcome E. Let τ be the patient interarrival time. Then the delayed outcome issue arises if $\tau < U_T$ or $\tau < U_E$. Figure 1 provides an example using the toxicity outcome. At the time τ a dose is assigned to a newly accrued cohort of patients, the toxicity outcomes of patients 2 and 3 are delayed. Although patient 2 has not experienced toxicity at time τ, he/she would experience toxicity later in $(\tau, U_T]$. At 3τ, we observe $T = 1$ for the first and second patients and $T = 0$ for the third patient.

From the missing data prospective, the delayed outcomes are nonignorable and the naive approach of discarding the missing part is problematic [44]. To handle the delayed outcomes, we first define V_T and V_E as the actual follow-up time of toxicity and efficacy for a patient

at the moment of decision making for dose assignment. It then follows that $T = 0$ only if $V_T = U_T$, which means that patient has been completely followed and no DLT is observed. Similarly, $E = 0$ only if $V_E = U_E$. let X_T and X_E be the time to toxicity and efficacy for a patient. The toxicity outcome is delayed only if $X_T > V_T$ and $V_T < U_T$, and the efficacy outcome is delayed only if $X_E > V_E$ and $V_E < U_E$. By using the weighted likelihood function method [45] and the Taylor expansion approximation [46], we approximate the DLT probability $\Pr(X_T > V_T|D_j)$ as

$$\begin{aligned}\Pr(X_T > V_T|D_j) &= \Pr(X_T > V_T, X_T > U_T|D_j) + \Pr(X_T > V_T, X_T \leq U_T|D_j) \\ &\approx 1 - p_{T,j} + p_{T,j}\frac{U_T - V_T}{U_T} = 1 - \frac{V_T}{U_T}p_{T,j} \approx (1 - p_{T,j})^{\frac{V_T}{U_T}}. \qquad (4)\end{aligned}$$

We assume that there are a total of n_{1j} patients being treated and completely followed at dose D_j. Among them, r_{1j} patients reported DLT. We also assume that there are a total of n_{2j} pending patients at dose D_j with their actual follow-up time for toxicity listed as $\{v_{j,1}, \cdots, v_{j,n_{2j}}\}$. Based on formula (4), the likelihood function for toxicity can be expressed as

$$L_T = \prod_{j=1}^{J}(p_{T,j})^{r_{1,j}}(1 - p_{T,j})^{n_{1j} - r_{1j} + \sum_{l=1}^{n_{2j}} \frac{v_{j,l}}{U_T}},$$

which can be treated as a series of binomial-like distributions. Noticing that because of the delayed outcomes, the effective sample size (ESS) has changed from $n_{1j}+n_{2j}$ to $n_{1j}+\sum_{l=1}^{n_{2j}} \frac{v_{j,l}}{U_T}$ to accommodate the incomplete follow-up time. The likelihood function for efficacy, L_E, can be derived using the same binomial approximation modeling approach. Then, all the aforementioned methods, such as the admissible set and modified isotonic regression, are directly applicable for these binomial-like distributions. Therefore, in the presence of delayed outcome issue, rather than suspending the trial for the pending patients, we can use this binomial approximation approach to derive the likelihood functions and then apply the mISO

design subsequently, which substantially shortens the trial duration, reduces administrative burden and saves resource. In addition, to strengthen the trial ethics, we require that dose assignment is not allowed until at least 50% of patients at the current dose level have completed both the DLT and efficacy assessment. Because this extension is based on the binomial distribution approximation, we refer to it as the mISO-B design.

We use Figure 1 again as an example to illustrate the application of the binomial distribution approximation. At time τ, patient 1 has experienced DLT, whereas the toxicity outcomes for patients 2 and 3 are missing with $v_2 = v_3 = \tau$ and $U_T = 3\tau$. Therefore, the approximated likelihood function accommodating delayed outcomes is $p(1-p)^{\frac{2}{3}}$ with the DLT estimate $\hat{p} = \frac{3}{5}$. Then, at time 2τ, patient 2 has experienced DLT and patient 3 is still at-risk, so we have updated v value as $v_3 = 2\tau$ and the likelihood changes to $p^2(1-p)^{\frac{2}{3}}$. The new DLT estimate is $\hat{p} = \frac{3}{4}$. Finally, at time 3τ when all the patients have completed their follow-up, the likelihood function with complete observation is $p^2(1-p)$, resulting in the final DLT estimate as $\hat{p} = \frac{2}{3}$.

5 Simulation studies

We conducted comprehensive simulation studies to investigate the operating characteristics of the mISO design. We considered a phase I/II trial with 6 doses and a maximum sample size of 60 in cohorts of size 3. We investigated the commonly used highest acceptable DLT rate of $\phi_T = 0.3$ and the lowest acceptable efficacy rate of $\phi_E = 0.5$. We set $\alpha_E = \beta_E = \alpha_T = \beta_T = 0.5$ in the beta-binomial model to specify non-informative priors. We also specified the cut-off probability values $\mu_T = 0.9$ and $\mu_E = 0.85$, which were determined through preliminary simulation studies for Bayesian design parameters calibration. We compared the mISO design with the non-parametric original isotonic regression-based design (ISO) we developed previously [18], two parametric model-based phase I/II designs EffTox and MTA-

RA [14, 26], and the conventional phase I followed by phase II design (CONV). The EffTox design is the first parametric model-based phase I/II design that uses a complicated six-parameter model to characterize the dose-response curves. The MTA-RA design also relies on the plateaued dose-efficacy curve. The main differences between the MTA-RA and mISO designs are that the MTA-RA design uses a logistic regression model for the dose-toxicity curve and a change point model for the dose-efficacy curve and determines the location of the plateau through a computational intensive Bayesian model averaging (BMA) approach. In contrast, the mISO design is parametric model-free and determines the plateau's location through the fast and standard AIC criterion, which can be easily obtained from any popular statistical software (e.g., SAS, R, SPSS, etc.). The CONV design is conventionally used in clinical practice, which uses a 3+3 design to find the MTD based on toxicity and then allocates more patients to the identified MTD to assess the efficacy further.

We first investigated the setting where the toxicity and efficacy outcomes can be immediately ascertained without the happening of delayed outcomes. Table 1 showed the simulation results, including the OBD selection percentage, the percentage of patients treated at each dose, and the sample size, which averaged over 10,000 simulated trials. We considered 6 scenarios representing a variety of dose-response curves. Under each scenario, the first and second rows represent the true underlying efficacy and toxicity rates we used to generate data. The OBD selection percentage under the dose level "0" indicates the percentage of early terminating the trial with no OBD identified. The numbers in bold were the correct OBD selection percentages for different designs.

In Scenario 1, the efficacy rate remained constant, so dose D_1 was the true OBD. The performance of the proposed mISO design was overwhelming, with a true OBD selection percentage of 82.6%, which was 45.9%, 22.7%, 55.8%, and 71.2% higher than that under the ISO, MTA-RA, EffTox, and CONV designs. It also yielded the highest percentage of 55.9% by assigning patients at the true OBD, with a percentage improvement ranging from 14.3%

to 35.5%, compared with other designs. In Scenarios 2 and 3, the OBD was located at dose D_2 and D_3. The mISO design was still the best, with the true OBD selection percentages of 53.7% and 54.7%.

In Scenarios 4 and 5, the OBD was located at dose D_4 and D_5. The mISO and MTA-RA were the top two designs, and the mISO consistently outperformed the MTA-RA in both OBD selection and patients allocation. Scenario 6 represents the "null" case where all the doses are either overly toxic or less efficacious. Because the mISO design has been equipped with an admissible set, when that set is empty indicating no acceptable doses, the mISO design can early terminate the trial. As expected, the mISO design yields a 100% probability of early terminating the trial in Scenario 6. The other designs also had a high percentage of about 90% to terminate the trial except for the ISO design because this design was not equipped with an early termination rule. Besides, as a consequence of early trial termination, the mISO, MTA-RA, and EffTox designs only use about one-third of the maximum sample size, which enhance the individual ethics of the trial. In summary, the mISO design was the best design for both the OBD selection and patients' allocation and should be recommended in practice.

We then investigated the delayed outcomes setting. Under this setting, the patient arrival time was simulated from Uniform distribution with the accrual rate 3 patients per month. The toxicity and efficacy assessment windows were 3 months. The time-to-efficacy and time-to-toxicity outcomes were simulated from Weibull distribution by controlling that 50% of the efficacy and toxicity occurred in the latter half of the assessment window $[1.5, 3]$. We compared the proposed mISO-B design with the conventional mISO-S design, which suspends the trial and defers the interim analysis if any pending patients exist.

Table 2 summarized the results. We reported the same operating characteristics and used the same DLT and toxicity rates for all the 6 scenarios, as in Table 1. Besides, we also reported the trial duration under each design. The mISO-B yielded slightly lower true OBD

selection and patients' allocation percentages than the conventional mISO-S design. As a reward, the mISO-B design almost halved the trial duration. For example, in Scenario 2, the mISO design reported a true OBD selection percentage of 52.8% and a patient's allocation at OBD percentage of 34.7%, which were only 1.9% and 1.6% lower than the counterparts under the mISO-S design. However, the mISO-B design substantially shortened the trial duration from 70.5 months to 39.5 months. The same pattern was also identified in other scenarios. Therefore, the proposed mISO-B is preferable to the conventional mISO-S design by considering all the design metrics including OBD selection, patients allocation and trial duration.

We conducted additional sensitivity analyses to investigate the robustness of the proposed mISO and mISO-B designs. First of all, in Table S1 of the online supplementary materials, we considered the umbrella-shape dose-efficacy response curves. The mISO design still reported desirable operating characteristics and yielded comparable performance compared with the ISO and MTA-RA designs. In Table S2, we varied the cohorts size of the mISO design and considered 2 patients per cohort and 5 patients per cohort, in addition to the 3 patients per cohort setting used in Table 1. As shown in Table S2, the default 3 patients per cohort setting yielded the best overall performances and we recommend this setting to be used in practice. In Figure S1 of the online supplementary materials, we reported the correct OBD selection percentages of the mISO design with different model selection criteria. We considered three criteria to determine the location of the plateau: the AIC used in this paper, the BIC, and the sum squared error (SSE) used in the original ISO design. Based on Figure S1, the AIC and BIC criteria yielded comparable performances and beat the SSE. Figures S2 to S4 also focused on the true OBD selection percentages but were depicted to evaluate the mISO-B design. Specifically, in Figure S2, we considered different patient arrival time distributions (Exponential and Uniform). In Figure S3, we varied the patient enrollment rate (2 patients/month, 3 patients/month, and 4 patients/month). In Figure S4,

we used different distributions to generate the time-to-toxicity and time-to-efficacy events (Log-logistic, Uniform, and Weibull). The corresponding sensitivity analysis for patients' allocation percentages at the OBD was also provided in the online supplementary materials (Figure S5 to Figure S8). We concluded that the mISO-B design is not vulnerable to different patient arrival distributions, patient enrollment rates, and time-to-event distributions based on all the simulation results.

6 Trial implementation

We carry out a hypothetical phase I/II clinical trial with the same setting as what is described in Table 1, with the exception that the hypothetical trial comprises five doses and a maximum sample size of 18. Figure 2 shows an illustrative trial example by comparing mISO-S to mISO-B, based on the patient-level data in Table 3. We can see that mISO-B shortens trial duration while having similar decisions to mISO-S.

For the mISO-B design, on day 101, when two patients of the first cohort complete endpoint assessment (e.g., $> 50\%$ of patients do not have pending data), we escalate the dose and treat the next cohort of patients at dose 2 as dose 1 is safe. If mISO-S were used, we would wait for 10 more days to observe all outcomes for the first cohort before enrolling the second cohort. With the mISO-B design, dose escalation continues until the fifth cohort. On day 455, patients 13 and 15 completed their assessment, and we reached the highest dose level. Since doses 2 and 3 are safe and efficacious, and OBD is identified at dose 2, we de-escalate back to dose 4 to treat the last cohort of patients starting on day 456. If mISO-S were used, it would take 546 days to get to this point. With mISO-B, on day 566, all patients complete the efficacy and toxicity assessment. The trial duration would be 656 days with the use of mISO-S. At the end of the trial, using either the mISO-B or mISO-S design, we found doses 2 and 3 are safe and efficacious, and dose 2 is estimated to be the

beginning location of the plateau and thus declared as the OBD.

7 Software

The R codes for implementation are available at https://github.com/yongzang2020/mISO. A Rode illustration and examples are provided in the online supplementary materials.

8 Conclusion

This paper proposes the mISO design, a transparent and efficient phase I/II clinical trial design, to identify the optimal biological dose for molecularly targeted agents and immunotherapy. The mISO design makes no parametric assumptions on the dose-response relationship and therefore yields desirable performances under any clinically meaningful dose-response curves. The concise, clinically interpretable model expression and dose-finding algorithm make the proposed designs highly translational from the statistical community to the clinical community. The original mISO design is only applicable for immediately ascertainable outcomes. Otherwise, a trial suspension is required due to delayed outcomes. To solve this problem, we extend the mISO design by proposing the mISO-B design to handle the delayed outcomes. The idea is to incorporate each pending patient's information by their remaining follow-up time and then use the binomial distribution approximation approach. Simulation studies show that the mISO and mISO-B designs have good operating characteristics for finding the optimal biological dose.

One interesting extension of the proposed designs is considering the optimal dynamic treatment regime that can adapt the same patient's dose assignment among different treatment cycles. In addition to single-agent trials, molecularly targeted agents and immunotherapy are often used together with other agents. It is of intense interest to extend the mISO and

mISO-B designs to identify the optimal dose combination for phase I/II drug-combination trials. We assume population homogeneity for all the subjects in the trial. However, an increased understanding of population heterogeneity of cancer has already brought us to the era of personalized medicine, providing clinicians with an unbeatable opportunity to select individually tailored treatments considering each subject's variability. Therefore, it is of interest to extend the proposed designs to integrate the personalized information into the trial. In this paper, the delayed toxicity and efficacy outcomes are modeled separately. If the optimal biological dose is defined in term of utility function, then the correlations of these delayed outcomes need to considered, and the conditional weighted likelihood method can be used under this setting [47]. Furthermore, if the occurrence of one endpoint will terminate the follow-up of the other one, then the statistical models handling competing-risks and semi competing-risks survival outcomes should be applied [48, 49, 50].

Acknowledgment

The authors thank the associate editor and referee for their valuable comments. Yong Zang's research is partially supported by NIH/NCI grants P30 CA082709, R21 CA264257, and the RalphW. and Grace M. Showalter Research Trust award. Yi Zhao's research is partially supported by NIH/NHLBI grant P01 HL158507.

Figure 1: Illustration of delayed outcomes mechanism. The horizontal line segment represents the follow-up. The toxicity outcome is indicated by an asterisk and triangles represent the arrival time of a new cohort.

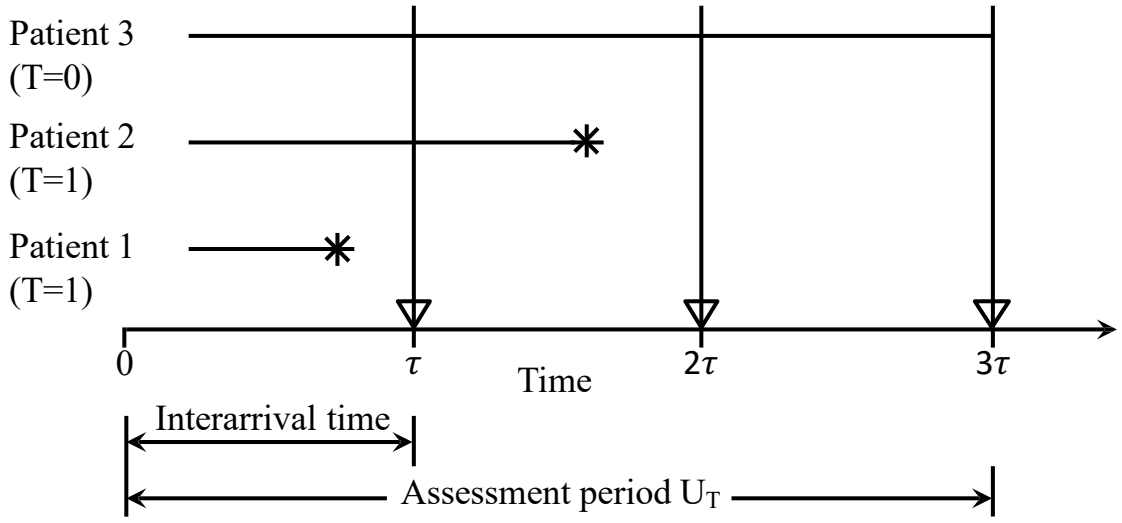

Figure 2: Trial illustration by applying the mISO-B design and the mISO-S design

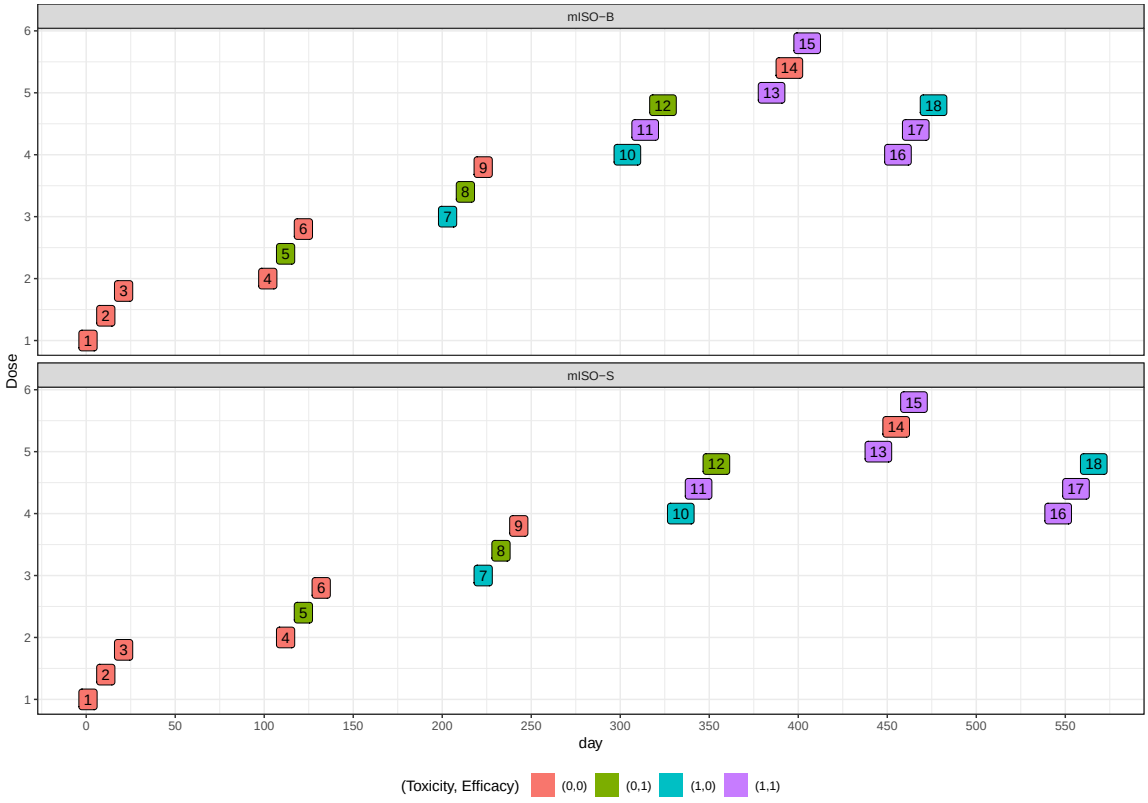

Table 1: The operating characteristics of the mISO, ISO, MTA-RA, EffTox and CONV designs including percentages of OBD selection and patients' allocation and number of patients in the trial, based on 10,000 simulated trials.

Design		Dose level							# of patients
		0	1	2	3	4	5	6	
				Scenario 1					
	True efficacy		0.8	0.8	0.8	0.8	0.8	0.8	
	True toxicity		0.03	0.1	0.2	0.3	0.4	0.5	
mISO	Selection (%)	1.4	**82.6**	12.2	2.6	1.0	0.1	0.0	
	Patient(%)		55.9	16.8	11.0	8.4	5.7	2.3	59.4
ISO	Selection (%)	0.0	**36.7**	28.7	16.9	10.7	6.3	0.7	
	Patient(%)		37.3	29.1	17.2	9.3	5.0	2.0	60.0
MTA-RA	Selection (%)	0.0	**59.9**	21.2	10.5	5.8	1.7	0.8	
	Patient(%)		41.6	22.4	16.2	12.5	5.9	1.3	60.0
EffTox	Selection (%)	0.0	**26.8**	52.6	16.6	3.0	1.0	0.0	
	Patient(%)		25.2	53.3	17.0	3.3	0.8	0.3	60.0
CONV	Selection (%)	6.6	**11.4**	29.8	31.9	15.5	4.4	0.3	
	Patient(%)		20.4	31.7	29.2	14.1	4.1	0.5	60.0
				Scenario 2					
	True efficacy		0.4	0.6	0.6	0.6	0.6	0.6	
	True toxicity		0.03	0.1	0.2	0.3	0.4	0.5	
mISO	Selection (%)	14.4	14.8	**53.7**	10.3	5.1	1.6	0.1	
	Patient(%)		25.3	36.4	15.1	11.8	8.0	3.4	54.7
ISO	Selection (%)	0.0	13.1	**41.4**	27.4	13.6	4.0	0.4	
	Patient(%)		18.2	37.4	25.2	12.8	4.9	1.5	60.0
MTA-RA	Selection (%)	2.7	10.2	**48.4**	23.2	11.6	3.4	0.4	
	Patient(%)		17.7	31.6	23.1	16.9	7.9	2.7	59.3
EffTox	Selection (%)	12.3	4.5	**24.0**	23.4	18.8	6.4	10.7	
	Patient(%)		8.4	20.2	21.3	20.8	15.4	14.0	55.2
CONV	Selection (%)	11.2	9.1	**28.4**	30.0	16.7	4.2	0.3	
	Patient(%)		20.1	31.9	28.5	14.9	4.2	0.5	60.0
				Scenario 3					
	True efficacy		0.2	0.4	0.6	0.6	0.6	0.6	
	True toxicity		0.03	0.1	0.15	0.3	0.4	0.5	
mISO	Selection (%)	19.9	0.3	15.7	**54.7**	7.6	1.6	0.1	
	Patient(%)		11.5	26.5	36.7	13.4	8.4	3.5	52.7
ISO	Selection (%)	0.0	11.9	14.2	**44.0**	22.8	6.5	0.5	
	Patient(%)		16.4	18.6	34.5	19.6	8.4	2.5	60.0
MTA-RA	Selection (%)	9.1	1.1	8.3	**43.1**	28.8	8.4	1.2	
	Patient(%)		10.5	17.4	29.3	25.5	12.8	4.5	57.7
EffTox	Selection (%)	12.1	0.6	0.6	**25.1**	31.4	14.8	15.6	
	Patient(%)		6.5	7.4	16.8	26.9	19.1	23.3	55.4
CONV	Selection (%)	18.1	5.9	16.8	**34.4**	18.6	5.9	0.3	
	Patient(%)		20.3	23.8	32.8	17.2	5.3	0.6	60.0

Table 1.(Continued)

Design		Dose level 0	1	2	3	4	5	6	# of patients
				Scenario 4					
	True efficacy		0.1	0.2	0.4	0.6	0.6	0.6	
	True toxicity		0.03	0.1	0.15	0.18	0.4	0.5	
mISO	Selection (%)	26.7	0.0	0.2	17.5	**51.8**	3.5	0.3	
	Patient(%)		9.8	13.1	27.7	34.8	10.6	4.1	49.9
ISO	Selection (%)	0.0	15.6	13.8	17.5	**39.2**	13.4	0.5	
	Patient(%)		18.2	14.0	12.4	34.5	15.9	5.0	60.0
MTA-RA	Selection (%)	19.1	1.4	1.6	7.2	**43.4**	23.6	3.7	
	Patient(%)		8.4	9.6	18.7	32.4	23.0	7.7	54.6
EffTox	Selection (%)	13.3	0.2	0.0	1.6	**28.1**	27.7	29.2	
	Patient(%)		6.5	6.6	8.1	24.5	27.0	27.4	54.9
CONV	Selection (%)	30.0	2.2	10.6	17.4	**30.0**	9.5	0.3	
	Patient(%)		20.4	23.3	21.4	25.2	8.8	0.9	60.0
				Scenario 5					
	True efficacy		0.1	0.2	0.3	0.4	0.75	0.75	
	True toxicity		0.03	0.08	0.1	0.15	0.2	0.5	
mISO	Selection (%)	20.8	0.0	0.1	2.2	12.8	**63.0**	1.0	
	Patient(%)		8.5	10.1	14.5	21.3	40.4	5.2	52.3
ISO	Selection (%)	0.0	15.6	18.6	14.7	8.4	**41.5**	1.3	
	Patient(%)		19.4	21.3	17.0	9.7	26.2	6.4	60.0
MTA-RA	Selection (%)	15.2	1.0	1.0	1.3	4.6	**58.7**	18.2	
	Patient(%)		6.7	6.7	14.8	17.5	30.3	24.1	54.7
EffTox	Selection (%)	12.3	4.5	10.7	23.4	18.8	**24.0**	6.4	
	Patient(%)		8.4	14.0	21.3	20.8	15.4	20.2	55.6
CONV	Selection (%)	30.3	1.6	3.7	11.2	20.4	**30.8**	1.8	
	Patient(%)		18.3	17.2	20.1	18.9	22.1	3.4	60.0
				Scenario 6					
	True efficacy		0.05	0.1	0.12	0.15	0.18	0.2	
	True toxicity		0.1	0.25	0.4	0.5	0.55	0.65	
mISO	Selection (%)	**100.0**	0.0	0.0	0.0	0.0	0.0	0.0	
	Patient(%)		29.9	29.5	22.0	11.8	5.2	1.6	14.2
ISO	Selection (%)	**0.0**	25.9	50.1	18.1	4.9	1.0	0.0	
	Patient(%)		24.5	38.7	22.0	9.9	3.7	1.1	60.0
MTA-RA	Selection (%)	**93.9**	0.2	5.9	0.0	0.0	0.0	0.0	
	Patient(%)		3.3	96.7	0.0	0.0	0.0	0.0	20.1
EffTox	Selection (%)	**95.0**	0.0	0.1	0.1	1.1	1.4	2.3	
	Patient(%)		18.1	18.0	17.6	16.6	16.3	13.4	19.1
CONV	Selection (%)	**89.3**	3.9	4.5	1.9	0.3	0.0	0.0	
	Patient(%)		55.2	33.4	9.9	1.4	0.1	0.0	60.0

Table 2: The operating characteristics of the mISO-B and mISO-S designs with delayed outcomes, including percentages of OBD selection and patients' allocation, number of patients in the trial and trial duration, based on 10,000 simulated trials.

Design		Dose level 0	1	2	3	4	5	6	# of patients	Duration (in months)
					Scenario 1					
	True efficacy		0.8	0.8	0.8	0.8	0.8	0.8		
	True toxicity		0.03	0.1	0.2	0.3	0.4	0.5		
mISO-B	Selection (%)	1.3	**83.9**	10.8	2.8	1.0	0.2	0.0		
	Patient(%)		55.0	16.7	11.1	8.6	6.1	2.4	59.5	42.5
mISO-S	Selection (%)	1.5	**82.1**	12.4	2.6	1.1	0.2	0.0		
	Patient(%)		55.4	16.8	11.0	8.6	5.8	2.3	59.4	76.2
					Scenario 2					
	True efficacy		0.4	0.6	0.6	0.6	0.6	0.6		
	True toxicity		0.03	0.1	0.2	0.3	0.4	0.5		
mISO-B	Selection (%)	15.3	16.8	**52.8**	9.3	4.3	1.5	0.1		
	Patient(%)		26.8	34.7	15.1	11.5	8.3	3.5	54.6	39.5
mISO-S	Selection (%)	13.7	14.5	**54.7**	10.8	4.8	1.3	0.1		
	Patient(%)		25.3	36.3	15.5	11.7	7.9	3.3	54.9	70.5
					Scenario 3					
	True efficacy		0.2	0.4	0.6	0.6	0.6	0.6		
	True toxicity		0.03	0.1	0.15	0.3	0.4	0.5		
mISO-B	Selection (%)	20.7	0.5	18.2	**51.8**	7.1	1.6	0.1		
	Patient(%)		11.8	28.3	34.3	13.5	8.6	3.5	52.7	38.3
mISO-S	Selection (%)	19.4	0.3	16.3	**54.6**	7.8	1.4	0.1		
	Patient(%)		11.6	27.3	36.4	13.2	8.2	3.3	52.8	67.8
					Scenario 4					
	True efficacy		0.1	0.2	0.4	0.6	0.6	0.6		
	True toxicity		0.03	0.1	0.15	0.18	0.4	0.5		
mISO-B	Selection (%)	27.7	0.0	0.2	18.8	**49.8**	3.0	0.4		
	Patient(%)		9.9	13.2	28.5	33.1	10.9	4.4	49.9	36.7
mISO-S	Selection (%)	27.9	0.1	0.3	16.4	**51.5**	3.4	0.4		
	Patient(%)		10.0	13.2	27.3	34.6	10.7	4.2	49.5	63.6
					Scenario 5					
	True efficacy		0.1	0.2	0.3	0.4	0.75	0.75		
	True toxicity		0.03	0.08	0.1	0.15	0.2	0.5		
mISO-B	Selection (%)	20.1	0.0	0.3	2.6	13.3	**62.1**	1.6		
	Patient(%)		8.3	10.0	14.8	22.7	38.6	5.6	52.7	38.7
mISO-S	Selection (%)	20.3	0.0	0.2	2.3	13.2	**63.2**	0.9		
	Patient(%)		8.4	9.8	14.6	21.9	40.2	5.2	52.6	67.8
					Scenario 6					
	True efficacy		0.05	0.1	0.12	0.15	0.18	0.2		
	True toxicity		0.1	0.25	0.4	0.5	0.55	0.65		
mISO-B	Selection (%)	**100.0**	0.0	0.0	0.0	0.0	0.0	0.0		
	Patient(%)		29.5	29.6	22.1	12.0	5.1	1.6	14.4	15.8
mISO-S	Selection (%)	**100.0**	0.0	0.0	0.0	0.0	0.0	0.0		
	Patient(%)		30.1	29.8	22.1	11.7	4.9	1.5	14.0	18.3

Table 3: Hypothetical patient-level data for trial illustration

mISO-S							mISO-B						
ID	Dose	T	E	Day	V_T	V_E	ID	Dose	T	E	Day	V_T	V_E
1	1	0	0	1	-	-	1	1	0	0	1	-	-
2	1	0	0	11	-	-	2	1	0	0	11	-	-
3	1	0	0	21	-	-	3	1	0	0	21	-	-
4	2	0	0	112	-	-	4	2	0	0	102	-	-
5	2	0	1	122	-	50	5	2	0	1	112	-	50
6	2	0	0	132	-	-	6	2	0	0	122	-	-
7	3	1	0	223	40	-	7	3	1	0	203	40	-
8	3	0	1	233	-	60	8	3	0	1	213	-	60
9	3	0	0	243	-	-	9	3	0	0	223	-	-
10	4	1	0	334	30	-	10	4	1	0	304	30	-
11	4	1	1	344	55	70	11	4	1	1	314	55	70
12	4	0	1	354	-	45	12	4	0	1	324	-	45
13	5	1	1	445	60	40	13	5	1	1	385	60	40
14	5	0	0	455	-	-	14	5	0	0	395	-	-
15	5	1	1	465	50	30	15	5	1	1	405	50	30
16	4	1	1	546	20	80	16	4	1	1	456	20	80
17	4	1	1	556	45	50	17	4	1	1	466	45	50
18	4	1	0	566	70	-	18	4	1	0	476	70	-

Abbreviations: Day, day of enrollment; ID, patient ID. "-" indicates that the patient doesn't experience DLT or efficacy outcome at the end of the follow-up.

References

[1] Sledge GW. (2005) What is target therapy. *Journal of Clinical Oncology*, 23: 1614-1615.

[2] Wu HC, Chang DK, Huang CT. (2006) Targeted therapy for cancer. *Journal of Cancer Molecules*, 2: 57-66.

[3] Topalian SL, Weiner GJ, Pardoll, DM. (2011) Cancer immunotherapy comes of age. *Journal of Clinical Oncology*, 23: 4828-4836.

[4] Couzin-Frankel J. (2013) Cancer immunotherapy. *Science*, 324: 1432-1433.

[5] Makkouk A, Weiner GJ. (2015) Cancer immunotherapy and breaking immune tolerance: new approaches to an old challenge. *Cancer Research*, 75: 5-10.

[6] Syn NL, Teng MWL, Mok TSK, Soo RA. (2017) De-novo and acquired resistance to immune checkpoint targeting. *Lancet Oncology*, 18: 731-741.

[7] Korn EL. (2004) Nontoxicity endpoints in phase I trial designs for targeted, non-cytotoxic agents. *Journal of National Cancer Institution*, 96: 977-978.

[8] Parulekar WR, Eisenhauer EA. (2004) Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: Theory and practice. *Journal of National Cancer Institution*, 96: 990-997.

[9] Postel-Vinay S, Arkenau HT, Olmos D, et al. (2009) Clinical benefit in phase I trials of novel molecularly targeted agents: Does dose matter? *British Journal of Cancer*, 100: 1373-1378.

[10] Tourneau CL, Lee JJ, Siu LL. (2009) Dose escalation methods in phase I cancer clinical trials. *Journal of National Cancer Institution*, 101: 708-720.

[11] LoRusso PM, Boerner SA, Seymour L. (2010) An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. *Clinical Cancer Research*, 16: 1710-1718.

[12] Thall P, Russell K (1998) A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. Statistics in Medicine, 27: 4895-4913.

[13] Braun, TM. (2002) The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Controlled clinical trials*, 23: 240-256.

[14] Thall PF, Cook J. (2004) Dose-finding based on efficacy-toxicity tradeoffs. *Biometrics*, 60: 684-693.

[15] Yin G, Li Y, Ji Y. (2006) Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics*, 62: 777-784.

[16] Zhang W, Sargent DJ, Mandrekar SJ. (2006) An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine*, 25: 2365-2383.

[17] Houede N, Thall PF, Nguyen H, et al. (2010) Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics*, 66: 532-540.

[18] Zang Y, Lee JJ, Yuan Y. (2011) Adaptive designs for identifying optimal biological dose for molecularly targeted agents. *Clinical Trials*, 11: 319-327.

[19] Thall PF, Nguyen HQ, Braun TM, et al. (2013) Using joint utilities of the times to response and toxicity to adaptively optimize schedule-dose regimes. *Biometrics*, 69: 673-682.

[20] Thall PF, Nguyen HQ, Zohar S Maton P. (2014) Optimizing sedative dose in preterm infants undergoing treatment for respiratory distress syndrome. *Journal of the American Statistical Association* 109: 931-943.

[21] Liu S, Johnson, VE. (2016) A robust Bayesian dose-finding design for phase I/II clinical trials. *Biostatistics*, 17: 249-263.

[22] Yuan Y, Yin, G. (2011) Bayesian phase I/II drug-combination trial design in oncology. *Annals of Applied Statistics*, 5: 924-942.

[23] Yuan Y, Nguyen HQ, Thall PF. (2016) Bayesian Designs for Phase I-II Clinical Trials. Chapman & Hall/CRC Biostatistics Series.

[24] Zhang L, Yuan Y. (2016) A Practical Bayesian Design to Identify the Maximum Tolerated Dose Contour for Drug Combination Trials. *Statistics in Medicine*, 35: 4924-4936.

[25] Guo B, Yuan Y. (2017) Bayesian phase I/II biomarker-based dose finding for precision medicine with molecularly targeted agents. *Journal of the American Statistical Association*, 112: 508-520.

[26] Riviere MK, Yuan Y, Jourdan,JH, Dubois F, Zohar S. (2018) Phase I/II dose-finding design for molecularly targeted agent: Plateau determination using adaptive randomization. *Statistical Method in Medical Research*, 27: 466-479.

[27] Guo B, Park Y, Liu S. (2018) A utility-based Bayesian phase I–II design for immunotherapy trials with progression-free survival end point. *Journal of the Royal Statistical Society-Series C*, 68: 411-425.

[28] Liu S, Guo, B. and Yuan, Y. (2018) A Bayesian phase I/II design for immunotherapy trials. *Journal of the American Statistical Association*, 113: 1016-1027.

[29] Yin G, Yuan, Y. (2009) Bayesian Model Averaging Continual Reassessment Method in Phase I Clinical Trials. *Journal of the American Statistical Association*, 104: 954-968.

[30] Horn L, Infante JR, Reckamp KL, et al. (2018) Ensartinib (X-396) in ALK-Positive Non-Small Cell Lung Cancer: Results from a First-in-Human Phase I/II, Multicenter Study. *Clinical Cancer Research*, 24: 2771-2779.

[31] Laubach JP, Liu CJ, Raje NS, et al. (2019) A Phase I/II Study of Evofosfamide, A Hypoxia-activated Prodrug with or without Bortezomib in Subjects with Relapsed/Refractory Multiple Myeloma. *Clinical Cancer Research*, 25: 478-486.

[32] Schroeder MA, Fiala MA, Huselton E, et al. (2019) A Phase I/II Trial of Carfilzomib, Pegylated Liposomal Doxorubicin, and Dexamethasone for the Treatment of Relapsed/Refractory Multiple Myeloma. *Clinical Cancer Research* 25: 3776-3783.

[33] Lowery MA, Bradley M, Chou JF, et al. (2019) Binimetinib plus Gemcitabine and Cisplatin Phase I/II Trial in Patients with Advanced Biliary Cancers. *Clinical Cancer Research* 25: 937-945.

[34] McMichael EL, Benner B, Atwal LS, et al. (2019) A Phase I/II Trial of Cetuximabin Combination with Interleukin-12 Administered to Patients with Unresectable Primary or Recurrent Head and Neck Squamous Cell Carcinoma. *Clinical Cancer Research* 25: 4955-4965.

[35] Haas NB, Appleman LJ, Stein M, et al. (2019) Autophagy Inhibition to Augment mTOR Inhibition: a Phase I/II Trial of Everolimus and Hydroxychloroquine in Patients with Previously Treated Renal Cell Carcinoma. *Clinical Cancer Research* 25: 2080-2087.

[36] Slovin S, Hussain S, Saad F, et al. (2019) Pharmacodynamic and Clinical Results from

a Phase I/II Study of the HSP90 Inhibitor Onalespib in Combination with Abiraterone Acetate in Prostate Cancer. *Clinical Cancer Research* 25: 4624-4633.

[37] Advani AS, Cooper B, Visconte V, et al. (2019) A Phase I/II Trial of MEC (Mitoxantrone, Etoposide, Cytarabine) in Combination with Ixazomib for Relapsed Refractory Acute Myeloid Leukemia. *Clinical Cancer Research* 25: 4231-4237.

[38] Thomas RP, Nagpal S, Iv M, et al. (2019) Macrophage Exclusion after Radiation Therapy (MERT): A First in Human Phase I/II Trial using a CXCR4 Inhibitor in Glioblastoma. *Clinical Cancer Research* 25: 6948-6957.

[39] Agrawal S, Feng Y, Roy A, Kollia G, Lestini B (2016) Nivolumab dose selection: challenges, opportunities, and lessons learned for cancer immunotherapy. *Journal for ImmunoTherapy of Cancer* 4: 72.

[40] Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD (1972) Statistical Inference under Order Restrictions. *London, New York: John Wiley & Sons.*.

[41] Hunsberger S, Rubinstein LV, Dancey J, Korn EL (2005) Dose escalation trial designs based on a molecularly targeted endpoint. *Statistics in Medicine* 24: 2171–2181.

[42] Hoering A, LeBlanc M, Crowley J (2011) Seamless phase I-II trial design for assessing toxicity and efficacy for targeted agents. *Clinical Cancer Research* 17: 640–646.

[43] Ivanova A, Flournoy N (2009) Comparison of isotonic designs for dose-finding. *Statistics in Biopharmaceutical Research* 1: 101.

[44] Yuan Y, Yin G (2011) Robust EM continual reassessment method in oncology dose finding. *Journal of American Statistical Association* 106: 818-831.

[45] Cheung YK, Chappell R (2000) Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 56: 1177–1182.

[46] Lin R, Yuan Y (2020) Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. *Biostatistics* 21: 807–824.

[47] Zhang Y, Zang Y (2021) CWL: A conditional weighted likelihood method to account for the delayed joint toxicity-efficacy outcomes for phase I/II clinical trials. *Statistical Methods in Medical Research* 30: 892-903.

[48] Zhang Y, Cao S, Zhang C, Jin IH, Zang Y (2021) A Bayesian adaptive phase I/II clinical trial design with lat-onset competing risk outcomes. *Biometrics* 77: 796-808.

[49] Biard L, Lee SM, Cheng B (2021) Seamless phase I/II design for novel anticancer agents with competing disease progression. *Statistics in Medicine* 40: 4568-4581.

[50] Zhang Y, Guo B, Cao S, Zhang C, Zang Y (2022) SCI: A Bayesian adaptive phase I/II dose-finding design accounting for semi-competing risks outcomes for immunotherapy trials. *Pharmaceutical Statistics* 21: 960-973.