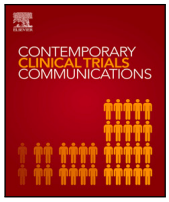
Contemporary Clinical Trials Communications 30 (2022) 101005

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc



Research paper

Interval design to identify the optimal biological dose for immunotherapy

Yeonhee Park

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, United States of America

ARTICLE INFO

Keywords:
Dose-finding
Immunotherapy
Interval design
Optimal biological dose

ABSTRACT

Immunotherapeutics have revolutionized the treatment of metastatic cancers and are expected to play an increasingly prominent role in the treatment of cancer patients. Recent advances in checkpoint inhibition show promising early results in a number of malignancies, and several treatments have been approved for use. However, the immunotherapeutic agents have been shown to have different mechanisms of antitumor activity from cytotoxic agents, and many limitations and challenges encountered in the traditional paradigm were recently pointed out for immunotherapy. I propose a desirability-based method to determine the optimal biological dose of immunotherapeutics by effectively using toxicity, immune response, and tumor response. Moreover, a new dose allocation algorithm of interval designs is proposed to incorporate immune response in addition to toxicity and tumor response. Simulation studies show that the proposed design has desirable operating characteristics compared to existing dose-finding designs. It also inherits the strengths of interval designs for dose-finding trials, yielding good performance with ease of implementation.

1. Introduction

Immunotherapeutics have revolutionized the treatment of metastatic cancers and are expected to play an increasingly dominant role in the treatment of cancer patients. Recent advances in checkpoint inhibition show promising early results in a number of malignancies [1]. Several treatments have been approved for use (e.g., ipilimumab, pembrolizumab, nivolumab) and many new immunotherapeutics are now being investigated in clinical trials (e.g., as of July 31, 2022, I found 1511 recruiting interventional studies for cancer immunotherapy in ClinicalTrials.gov). While traditional clinical trials remain the industry standard for dose-finding and testing safety and efficacy, many flaws in this paradigm were recently pointed out [2–4]. Specifically, immunotherapeutic agents have been revealed to have different toxicity profiles and mechanisms of antitumor activity from cytotoxic agents.

As an effective way for dose-finding to address the challenge, phase I/II designs are developed with a model-based algorithm [5–7]. The model-based phase I/II clinical trial designs are more accurate and robust across studies for immunotherapeutics because they incorporate immune response as well as toxicity and tumor response. However, they are generally difficult for clinicians to understand and implement due to their conceptual and computational complexity. In addition, statistical analyses are repeatedly required to be implemented after each dose cohort or after each interim decision point. Due to the difficulties of the model-based phase I/II clinical trial designs, clinical practice has been dominated by designs that are simpler and more straightforward to implement. For example, recently proposed phase I/II interval designs [8,9] offer simpler dose allocation rule solutions. However, they are still limited when applied to immunotherapy, because they do not incorporate immune response measuring the biological efficacy of the agents in activating the immune system.

Motivated by these challenges of the dose-finding trials for immunotherapy, I propose an interval design based on toxicity, immune response, and tumor response (ITIT) to find the optimal biological dose (OBD) of immunotherapeutics. The ITIT effectively uses clinical outcomes for immunotherapy trials, such as toxicity, immune response, and tumor response, in order to reflect the unique features of immunotherapeutics. One of the advantages of ITIT is the simplicity to implement in practice based on the clinical outcomes. In this design, dose escalation/de-escalation is determined by comparison of the observed probability of toxicity, immune response, and tumor response with the prespecified boundaries. This does not require complicated computation, and the identified boundaries minimize the classification error rate. In addition, the use of immune response improves the dose allocation by assuming dose with immune response as positive responses.

Compared to existing dose-finding designs, the proposed design makes four new contributions. First, dose allocation algorithm incorporates three essential outcomes for immunotherapy trials, which is different from most existing phase I or phase I/II dose-finding designs based on toxicity only or toxicity and tumor response. It elaborates and improves the decision. Secondly, the proposed design investigates the

E-mail address: ypark56@wisc.edu.

https://doi.org/10.1016/j.conctc.2022.101005
Received 30 December 2021; Received in revised form 31 July 2022; Accepted 16 September 2022
Available online 24 September 2022

toxicity of the dose prior to clinical benefits and safely allocates doses, which is critical to ensure patients safety in early phase trials. Third, it proposes methods for a desirability-based selection of OBD with three clinical outcomes. This determines OBD as the tolerable dose with a high probability of immune response and overall response. Lastly, the proposed design is simple and straightforward to implement. It avoids conceptual and computational complexity while it considers the unique features of immunotherapeutics.

The rest of this paper is organized as follows. In Section 2, I propose an interval design and describe the selection of OBD. I evaluate the operating characteristics of the proposed design using simulation studies in Section 3 and provide concluding remarks in Section 4.

2. Methods

2.1. Interval design to identify the optimal biological dose based on toxicity, immune response, and tumor response (ITIT)

Consider a phase I/II immunotherapy trial with a total of J prespecified doses of a new drug under investigation. During the trial, toxicity, immune outcome, and tumor response are measured. Let Y_T denote an indicator of having experienced dose limiting toxicity. Let Y_I denote the immune outcome, for example, in NCT02523469 trial, the count of CD8+ T cells, CD4+ T cells, or NK cells were measured. Given a cutoff c_I, specified by clinicians such that $Y_I \geq c_I$ denotes the desirable immune response, the transformed immune outcome $\tilde{Y}_I$, which takes 1 if $Y_I \geq c_I$, and 0 otherwise, indicates the immune response such as the immune activation state or functionality of the immune cells. The tumor response is often assessed by the immune complete response (iCR), immune partial response (iPR), immune stable disease (iSD), immune unconfirmed progressive disease (iUPD) or immune confirmed progressive disease (iCPD) according to iRECIST criteria [10]. In practice, clinical researchers are interested in overall response, including iCR or iPR, to assess the efficacy of a drug, and I use the notation Y_E to denote the indicator of iCR or iPR if necessary. My approach is to incorporate all three clinical outcomes for immunotherapy in order to allocate the dose to the next patient(s) and find the recommended dose of the new drug, which is different from most existing phase I or phase I/II dose-finding designs based on toxicity only or toxicity/efficacy outcomes, respectively.

Let ϕ_T, ϕ_I, and ϕ_E denote the target rate for toxicity, immune response, and overall response, respectively, which are specified by physicians. Assume that patients are treated in cohorts, but allow the size of a cohort to vary from one cohort to another. Then, an interval design based on toxicity, immune response and tumor response (ITIT) for dose-finding trials is described as follows. Patients in the first cohort are treated at the lowest dose level. The design does not allow to skip the untested doses. For $j = 1, \ldots, J$, let $\hat{p}_{T,j}$, $\hat{p}_{I,j}$, and $\hat{p}_{E,j}$ denote the estimated probability for toxicity, immune response, and overall response, respectively, based on all the accumulative data on the current dose level j. Assume that at the current dose level j, a total of n_j patients have been treated, x_j of them have experienced toxicity, z_j of them have immune response, and y_j of them show iCR or iPR in tumor response. Based on the observed data, the probabilities for toxicity, immune response, and overall response are estimated by $\hat{p}_{T,j} = x_j/n_j$, $\hat{p}_{I,j} = z_j/n_j$, and $\hat{p}_{E,j} = y_j/n_j$, respectively. Let λ_1 and λ_2 denote the prespecified lower and upper boundaries, respectively, for the estimated toxicity probability satisfying $0 \leq \lambda_1 < \phi_T < \lambda_2 \leq 1$; let η denote the cutoff on the estimated immune response rate satisfying $0 \leq \eta < \phi_I < 1$; and let δ denote the prespecified boundary for the estimated overall response rate (ORR) satisfying $0 \leq \delta < \phi_E < 1$. The optimal values of λ_1, λ_2, η, and δ are identified in the sense that they minimize the dose misclassification probability (see the details in Appendix A). To assign a dose to the next cohort of patients,

1. if $\hat{p}_{T,j} \geq \lambda_2$, the current dose level j is deemed overly toxic and I de-escalate the dose level to $j-1$.
2. if $\lambda_1 < \hat{p}_{T,j} < \lambda_2$, the current dose level j is deemed close to the target toxicity rate, which may be risky to escalate to an overly toxic dose or de-escalate to the subtherapeutic dose. It is desirable to stay at the current dose and observe more data to justify the decision for dose allocation.
3. Otherwise, i.e., $\hat{p}_{T,j} \leq \lambda_1$, the current dose level j is regarded as being safe or overly safe to stay or escalate, respectively. Clinical outcomes, such as immune response and tumor response, are used to determine the dose allocation.
   (a) if $\hat{p}_{E,j} > \delta$, the current dose level is considered as the preferable dose level in terms of both toxicity and efficacy, and I retain the same dose level j.
   (b) Otherwise, i.e., $\hat{p}_{E,j} \leq \delta$, I look at immune outcome: the current dose can be regarded promising or effective when $\hat{p}_{I,j} > \eta$, and otherwise, it indicates the dose at which immune cells are less likely to be activated. Therefore, I stay at the current dose if $\hat{p}_{I,j} > \eta$, and otherwise I escalate the dose level to $j+1$.

The decision for dose escalation or de-escalation is made by comparing the estimated probabilities $\hat{p}_{T,j}, \hat{p}_{E,j}$, $\hat{p}_{I,j}$ with the boundaries λ_1, λ_2, δ, and η. The proposed dose allocation rule uses essential outcomes for immunotherapy trials while the decision order prioritizing the safety over other clinical outcomes strengthens the ethics of the proposed design. In early phase clinical trials, there is very limited understanding on the new agents, thus it is of paramount importance to ensure patients' safety. In addition, the proposed design elaborates the decision for immunotherapeutics by examining immune response and tumor response for safe doses. When a dose is deemed to be safe, the interval of the estimated ORR is divided into two subintervals $[0, \delta]$ and $(\delta, 1]$ to denote the dose as being less efficacious or more efficacious, respectively. Even a less efficacious dose (i.e., $\hat{p}_{T,j} \leq \lambda_1$ and $\hat{p}_{E,j} \leq \delta$) can be likely to yield a positive response when the immune cells are proliferated. Thus, I incorporate information from the immune response to decide the dose assignment with escalation or retainment when $\hat{p}_{I,j} \leq \eta$ or $\hat{p}_{I,j} > \eta$, respectively, for the safe but inefficacious dose. The current dose satisfying $\hat{p}_{T,j} \leq \lambda_1$, $\hat{p}_{E,j} \leq \delta$, and $\hat{p}_{I,j} \leq \eta$ is not worthy of staying but worthy of escalation to explore the doses. This decision is ethical and avoids trapping at the current dose while the statistical complication is minimized. Such sequential decisions can be summarized in Fig. 1. According to the dose allocation rule, the support of probabilities for toxicity, immune response, and overall response is partitioned into 12 regions according to the cutoffs λ_1, λ_2, η, and δ and the decisions (D/S/E) can be recorded on the regions. It follows the rationale of the dose allocation but simplify the procedure directly without multiple steps of the decision procedure.

The dose allocation rule can be modified to be less conservative. I considered several variants of the dose allocation rules for the sensitivity analysis in the simulation study. When $\hat{p}_{T,j} \leq \lambda_1$ and $\hat{p}_{E,j} \leq \delta$, I can generate a binary indicator with success probability $\hat{p}_{I,j}$ to escalate the dose. If a success is achieved, I retain the dose level j. Otherwise, I escalate the dose level to $j+1$. Alternatively, to avoid using random walk to allocate, the variants of the dose allocation rule fix escalation when (1) $\hat{p}_{T,j} \leq \lambda_1$ and $\hat{p}_{E,j} \leq \delta$, (2) $\hat{p}_{T,j} \leq \lambda_1$ except the subregion where $\hat{p}_{E,j} > \delta$, and $\hat{p}_{I,j} > \eta$, or (3) $\hat{p}_{T,j} \leq \lambda_1$.

2.2. Selection of optimal biological dose

While conventional dose-finding trials have pursued Maximum Tolerated Dose (MTD) as the objective, it is increasingly acknowledged that the objective for immunotherapies should be Optimal Biological Dose (OBD) to achieve the optimal treatment effect or risk–benefit tradeoff. At the end of the trial, accumulating data allow to estimate the probability of toxicity, immune response, and overall response for each dose, so that the biologic efficacy of the immunotherapeutic agent in activating the immune system and the efficacy in clinical

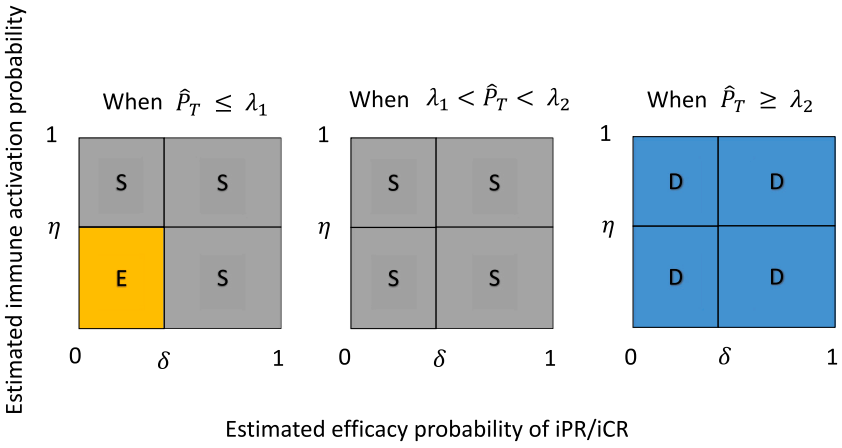


Fig. 1. Partitioned regions for dose allocation rule under the proposed design.

Table 1
Desirability score using probabilities p_T, p_I, and p_E.

	When $p_T \le \phi_T$			
	$p_E < 0.6\phi_E$	$0.6\phi_E \le p_E < 0.85\phi_E$	$0.85\phi_E \le p_E < \phi_E$	$p_E \ge \phi_E$
$p_I < 0.2\phi_I$	10	50	70	80
$0.2\phi_I \le p_I < 0.6\phi_I$	25	50	70	80
$0.6\phi_I \le p_I < \phi_I$	35	50	70	80
$p_I \ge \phi_I$	45	55	90	100
	When $p_T > \phi_T$			
	$p_E < 0.6\phi_E$	$0.6\phi_E \le p_E < 0.85\phi_E$	$0.85\phi_E \le p_E < \phi_E$	$p_E \ge \phi_E$
$p_I < 0.2\phi_I$	0	18	25	28
$0.2\phi_I \le p_I < 0.6\phi_I$	9	18	25	28
$0.6\phi_I \le p_I < \phi_I$	11	18	25	28
$p_I \ge \phi_I$	16	19	32	35

benefit can be investigated. I define OBD as the tolerable dose with a high probability of immune response and overall response, which effectively incorporates all essential outcomes from the immunotherapy trials. To define OBD, I describe the desirability of multiple outcomes x, z, and y to measure the risk–benefit tradeoff that underlies therapeutic decision making. It is the function of p_T, p_I, and p_E, which denote the true probability for toxicity, immune response, and overall response, respectively. I denote it by $U(p_T, p_I, p_E)$. To facilitate the elicitation, I fix the desirability of the most desirable outcome pair as $U(p_T, p_I, p_E) = 100$ and the desirability of the least desirable outcome pair as $U(p_T, p_I, p_E) = 0$, and then ask clinicians to use these two pairs as references to score the desirability of the other elementary outcome pairs $\{U(p_T, p_I, p_E)\}$ using the scale of $(0, 100)$. For example, the desirability scores can be elicited for immunotherapy trials in Table 1. I partitioned the probability of immune response into subintervals, $[0, 0.2\phi_I)$, $[0.2\phi_I, 0.6\phi_I)$, $[0.6\phi_I, \phi_I)$ (being far from target, less close to target, close to target, respectively), and $[\phi_I, 1]$ (reaching the target value), and partitioned the probability of the overall response into subintervals indicating being inefficacious, i.e., $[0, 0.6\phi_E)$, and being efficacious, $[0.6\phi_E, 0.85\phi_E)$, $[0.85\phi_E, \phi_E)$, and $[\phi_E, 1]$ (less likely, close to target, more likely, respectively). Those boundaries of subintervals are specified by considering 40% deviation from the target immune response and tumor response to decide high immune response and distinguish between inefficacious and efficacious doses. Since immune activation can be a positive response, which is not certainly confirmed, I elaborate on the region of $< 0.6\phi_I$ rather than $\ge 0.6\phi_I$. Since overall response is a primary measure of efficacy, I elaborate on the region of $\ge 0.6\phi_E$ to help distinguish the doses based on the estimated efficacy probability.

To safeguard patients from overly toxic doses, a set of acceptable doses for safety based on the estimated toxicity probability is defined by using the dose d^* whose toxicity probability is closest to the target toxicity rate. Dose level j is deemed acceptable for safety if $d_j \le d^*$. Given acceptable doses, I estimate the desirability $U(\hat{p}_{T,j}, \hat{p}_{I,j}, \hat{p}_{E,j})$ by plugging in the estimated probabilities $\hat{p}_T$, $\hat{p}_I$, and $\hat{p}_E$ based on the accumulating data for the dose level j. Then, the OBD is determined as the dose of level

$$\underset{j \in \{i : d_i \le d^*\}}{\text{argmax}} \; U(\hat{p}_{T,j}, \hat{p}_{I,j}, \hat{p}_{E,j}),$$

which corresponds to the most desirable dose with respect to the immune response and tumor response among the acceptably safe doses. The selected dose d^* is known to be the MTD, which phase I trials are commonly used for the recommended dose for phase II trials. In the case where d^* is larger than ϕ_T, the use of desirability function putting a penalty on overly toxic doses reduces the chance of misidentifying an overly toxic dose as OBD. Thus, the restriction defines the admissible set of doses to identify the OBD.

The desirability-based method above is very straightforward and easy to implement based on p_T, p_I, p_E. However, if the study has primarily objective to investigate mechanisms of antitumor activity for immunotherapeutic agents, joint probability models of essential outcomes could provide useful interpretations [5,7]. The desirability score can be built based on the outcomes $Y_T = a, \tilde{Y}_I = b, Y_E = c$, where $a, b, c = 0, 1$ instead of the probabilities p_T, p_I, p_E. The expected desirability is calculated based on the score and the joint probabilities $\Pr(Y_T = a, \tilde{Y}_I = b, Y_E = c)$ to choose OBD maximizing the expected desirability of dose.

2.3. Practical implementation

To implement the ITIT in practice, I need to specify design parameters $\phi_{T,1}, \phi_{T,2}$, $\phi_{E,1}$, and $\phi_{I,1}$, where $\phi_{T,1}$ denotes the highest toxicity probability that is deemed safe and $\phi_{T,2}$ denotes the lowest toxicity probability that is deemed overly toxic with $0 < \phi_{T,1} < \phi_T < \phi_{T,2} <$

1; $\phi_{E,1}$ denotes the highest ORR that is deemed least desirable with $0 < \phi_{E,1} < \phi_E < 1$; and $\phi_{I,1}$ denotes the highest immune response probability that is deemed less activated such that $0 < \phi_{I,1} < \phi_I < 1$. To specify interval boundaries λ_1 and λ_2, $\phi_{T,1} = 0.6\phi_T$ and $\phi_{T,2} = 1.4\phi_T$ are recommended for the default value to decide a maximum tolerated dose from the phase I trials [11]. The same approach deviating 40% from the target is used for tumor response and immune response, i.e., I recommend $\phi_{E,1} = 0.6\phi_E$ for the specification of δ and $\phi_{I,1} = 0.6\phi_I$ for the specification of η. Existing phase I/II interval designs suggest different design parameters, which are either closer to the target (e.g., $\phi_{T,1} = 0.75\phi_T$) or smaller than the recommended values (e.g., $\phi_{T,1} = 0.1\phi_T$). The best approach is to reflect the clinical need and practice for the specification of the parameters, and close collaboration with clinicians and preliminary simulations help elicitation of the parameters.

ITIT assumes that three outcomes are observed for the decision before the next new patient is accrued. However, immunotherapeutic drugs often have delayed toxicity outcomes. Moreover, patients' data can be pending because of the rapid accrual rate and different assessment windows. The responses can happen much later in which case it would delay the decision for dose escalation/de-escalation between cohorts in order to observe them. To take into account the timing of events for the binary outcomes, the imputation method for delayed outcomes can be used [12]. Briefly speaking, let Y_i denote a binary endpoint indicator for the ith patient treated at the current dose (e.g., Y_T, $\tilde{Y}_I$, or Y_E) and n denote the number of patients enrolled at the current dose. Suppose that some data are observed after the assessment completion while others are pending. Let O and M denote the sets of patients whose binary endpoint data are observed and pending, respectively. For $i \in M$, y_i is imputed by the expected value for the ith patient treated at the current dose with the follow-up time F_i, i.e., $E(y_i|T_i > F_i)$, where T_i denotes the time to the binary event. Yuan et al. [12] shows that $E(y_i|T_i > F_i)$ is approximated by $p(1 - F_i/W)/(1 - p)$, where W denotes the prespecified assessment window. Then, the estimated probability for the binary endpoint is calculated by

$$\hat{p} = \frac{\sum_{i\in O} y_i + \sum_{i\in M} y_i}{n} = \frac{\sum_{i\in O} y_i + \sum_{i\in M} \hat{E}(y_i|T_i > F_i)}{n},$$

where $\hat{E}(y_i|T_i > F_i)$ is obtained by replacing p with the posterior mean of p based on the observed data. This imputation approach provides the probabilities $\hat{p}_T, \hat{p}_I, \hat{p}_E$. Then, after the imputation, the probabilities are compared with the boundaries $\lambda_1, \lambda_2, \eta$, and δ to determine the dose allocation.

3. Simulation study

I evaluated the operating characteristics of the proposed design using simulations. I considered five doses, with a maximum sample size of 30 patients in a cohort size of 3. Followed by phase I dose-finding trials, the sample size of 30 was obtained by 6 times the number of doses, and the cohort size of 3 was used conventionally. Suppose that target rates for toxicity, immune response, and overall response were $\phi_T = 0.3$, $\phi_I = 0.5$, and $\phi_E = 0.7$, respectively. I set $\phi_{T,1} = 0.18$, $\phi_{T,2} = 0.42$, $\phi_{I,1} = 0.3$, and $\phi_{E,1} = 0.42$, which denote 40% deviation from the target rates. The equal prior probabilities were assigned throughout the simulation, and optimal boundaries were calculated by Eqs. (1), (2), and (3) of Appendix A: $\lambda_1 = 0.236, \lambda_2 = 0.359, \eta = 0.397$, and $\delta = 0.563$ were obtained. The boundaries were used to classify the dose in terms of toxicity, immune response, and overall response. The desirability score in Table 1 was used to make a decision for OBD based on clinical outcomes.

I compared the proposed design (ITIT) with three designs: a design utilizing efficacy and toxicity (denoted as STEIN design) and designs utilizing toxicity only (denoted as BOIN and 3+3 design). ITIT, STEIN, and BOIN were interval designs while 3+3 design was a rule-based (or algorithm-based) design. Even though 3+3 design was not an interval design, it was the most popular Phase I dose-finding design, which was dominant in practice, and I included the design for the comparison. The same values of design parameters as ITIT (e.g., $\phi_{T,1} = 0.18$, $\phi_{T,2} = 0.42$, or $\phi_{E,1} = 0.42$) were used for STEIN and BOIN to make the comparisons meaningful. In addition, STEIN and BOIN implemented a dose elimination rule eliminating the dose levels j and higher such that $\Pr(p_{T,j} > \phi_T|x_j, n_j) > 0.95$ (see the details in [9] and [11]). Note that Lin and Yin [9] shows that STEIN design generally outperforms the existing model-based methods for molecularly targeted agent in [13] and [14]. A rich body of literature on BOIN shows comparable performance with the Continual Reassessment Method in [15], which is the most popular model-based design for Phase I dose-finding trial [11,16,17].

Fig. 2 illustrated the scenarios in the simulation study and showed the curves between dose level and clinical outcomes. Those curves demonstrated various shapes and locations of the OBD. Scenario 1 had OBD at dose level 1, since all five doses were acceptably safe but both immune and tumor responses decreased in dose level. Scenario 2 also had OBD at dose level 1, since the first dose level is only one whose toxicity probability is smaller than or equal to the target toxicity rate. In scenario 3, all dose levels were acceptably safe but dose level 2 had the maximum probability of overall response. In addition, immune response probability increased with the dose level but almost plateaued at dose level 2. Thus, scenario 3 had OBD at level 2. In scenario 4, dose levels 1 and 2 had acceptable toxicity, but level 2 denoted OBD, because it had much higher probabilities of immune response and overall response than level 1. In scenario 5, the first four dose levels were acceptably safe, but both immune response and overall response had the maximum at level 3 with a ∩-shaped curve. Thus, level 3 denoted OBD. Scenario 6 had OBD at dose level 3, which had the maximum probabilities of immune response and overall response among the acceptably safe doses at levels 1–3. For similar reasons to scenario 6, scenarios 7 and 8 had OBD at dose level 4. In scenarios 9 and 10, all dose levels were safe but had the maximum efficacy at level 5. In particular, the toxicity probability curve in scenario 10 was flat over the dose levels, but immune response and tumor response increased with the dose level.

I simulated 10,000 trials and summarized in Table 2 the operating characteristics of ITIT, STEIN, BOIN, and 3+3 design including the selection percentage and the average number of patients treated at each dose. In scenario 1, OBD was the lowest dose, which was identified with a higher power under ITIT than other designs by using immune response and overall response effectively. The percentage of correct selection of OBD under ITIT was 24% higher than under the STEIN design and 78% higher than under the BOIN and 3+3 design. The number of patients treated at the OBD under ITIT was larger than with the other designs. In scenario 2, OBD was the lowest dose but the only level with a toxicity probability smaller than or equal to the target rate. For a higher probability of immune response than the target rate, the dose allocation rule of ITIT did not allow the dose to escalate. In addition, dose level 2 had a close toxicity probability to the target, and both the BOIN and 3+3 design did not distinguish the dose levels 1 and 2 well. Thus, OBD was chosen with 22–49% more under ITIT than under the other designs. In scenarios 3, 6, 8, by taking advantage of the immune response, ITIT distinguished dose levels with similar overall responses and yielded a higher percentage of selection at OBD. In scenarios 4, 5, 7, the selection percentage gain of ITIT was attained from effective use of immune response and tumor response compared to other designs. In scenarios 9 and 10, OBD was the highest dose, which was the same as MTD, and BOIN outperformed other designs. ITIT still worked well to have a larger percentage of correct selection than STEIN and 3+3 design. ITIT could be improved in this situation by using the modified dose allocation rule which was less conservative to explore higher doses with potentially higher efficacy when the current dose was safe or by specifying $\phi_{E,1} = 0.9\phi_E$ for the specification of δ. The numbers of patients treated above the optimal dose under ITIT were 3.97, 5.87, 3.77, 1.97, 0.51, 0.45, 0.18, 0.15, 0, and 0 for scenarios 1–10, respectively.

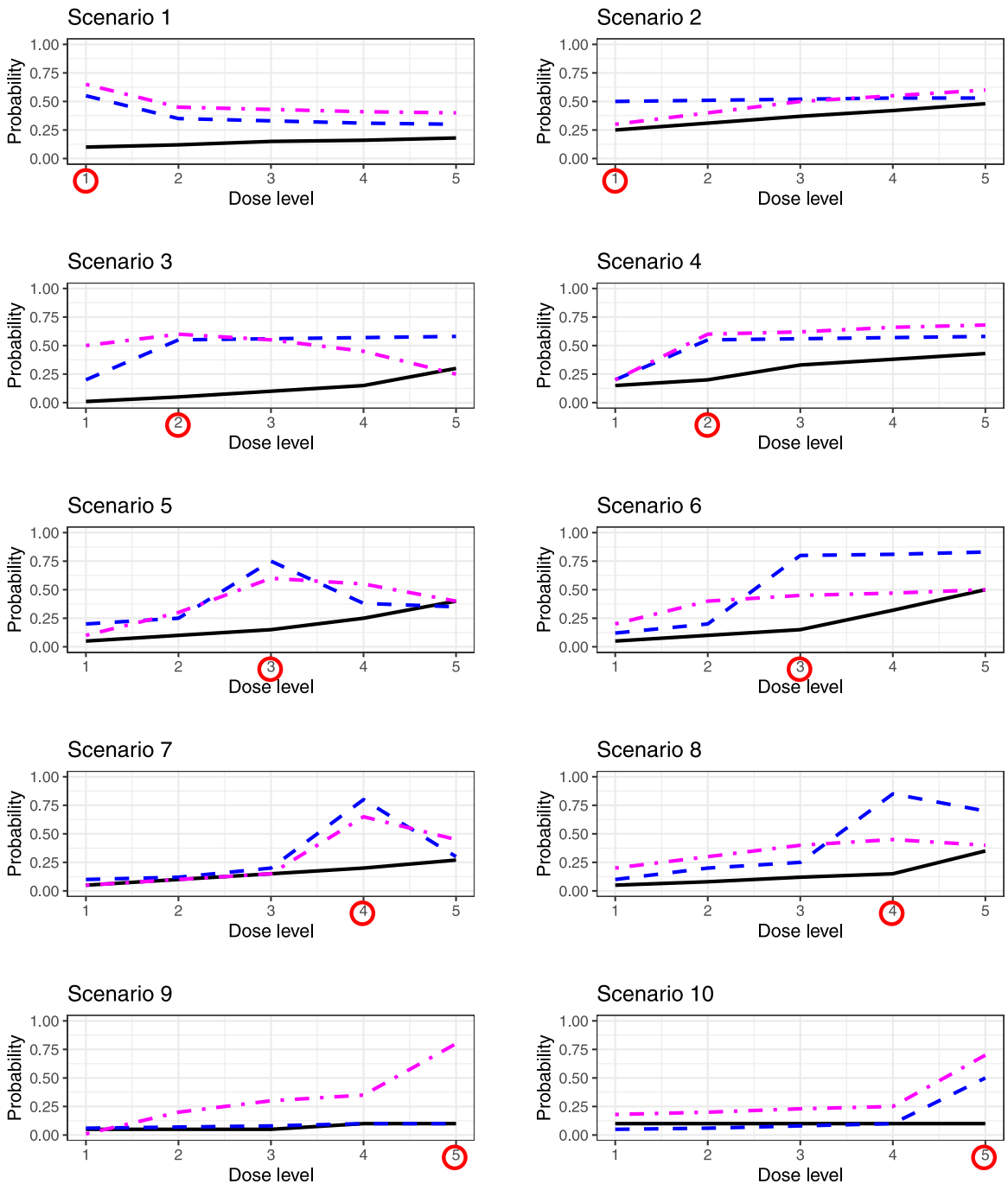


Fig. 2. Dose–response curves for the ten scenarios in the simulation study. The solid, dashed, and dotdash lines denote the toxicity, immune response, and overall response, respectively. The red circle around axis labels indicates the OBD.

It was not surprising to observe the gain of the proposed design in terms of the selection percentage and the number of patients treated at OBD, because the proposed design utilized more endpoints (i.e., all three available endpoints) to make the decision effectively. The simulation study aimed to bring the issues of using conventional dose-finding designs (BOIN or 3+3 design) for immunotherapy in practice and emphasize the appropriate decision for the recommended phase II dose in clinical practice. Since the BOIN and 3+3 design were more likely to select MTD based only on toxicity outcome, their performance in the selection percentage was not good when MTD and OBD were not equal. In particular, when OBD was far away from MTD, BOIN's performance was not good (e.g., scenarios 1–8) while it performed better when OBD and MTD were equal at the highest dose level (e.g., scenarios 9–10).

As for sensitivity analyses, I investigated the performance of the proposed design with three different desirability scores from Table 1. The different desirability scores and results were presented in Web Appendix A. Using more elaborated safe subintervals for the desirability showed similar operating characteristics while considering a few subintervals of immune/tumor responses for the desirability would not distinguish the dose desirability well based on the estimates. Also, the desirability score ignoring immune response did not work well when toxicity and efficacy profiles were not different in doses. I also evaluated the sensitivity of the ITIT design to the maximum sample size and target rates. The maximum sample size of 51 (i.e., 17 cohorts of size 3) was considered instead of a maximum sample size of 30. Web Table 5 showed that the ITIT design was rather robust to the maximum

Table 2
Simulation scenarios with true probability (p_T, p_I, p_E) and true desirability at each dose and simulation results for selection percentage (the number of patients treated) at each dose.

	Dose level				
	1	2	3	4	5
	Scenario 1				
p_T	0.1	0.12	0.15	0.16	0.18
p_I	0.55	0.35	0.33	0.31	0.3
p_E	0.65	0.45	0.43	0.41	0.4
Desirability	**90**	50	50	35	35
ITIT	**89.54(25.97)**	5.11(1.86)	2.71(1.13)	1.53(0.60)	0.86(0.39)
STEIN	**65.55(19.3)**	13.38(4.3)	10.02(3.0)	7.38(2.0)	3.39(1.2)
BOIN	**0.91(4.73)**	3.46(5.19)	9.30(5.60)	14.35(5.07)	71.67(9.34)
3 + 3	**11.94(6.25)**	15.03(6.12)	13.60(4.99)	14.61(4.08)	35.10(6.16)
	Scenario 2				
p_T	0.25	0.31	0.37	0.42	0.48
p_I	0.5	0.51	0.52	0.53	0.53
p_E	0.3	0.4	0.5	0.55	0.6
Desirability	**45**	16	19	19	32
ITIT	**79.42(22.38)**	10.65(5.05)	1.31(0.75)	0.13(0.06)	0.00(0.00)
STEIN	**57.07(16.0)**	25.14(8.3)	7.36(3.2)	1.24(0.6)	0.10(0.1)
BOIN	**36.07(13.63)**	31.04(8.75)	16.57(4.18)	6.45(1.38)	1.65(0.37)
3 + 3	**30.56(10.69)**	16.61(5.50)	6.50(2.13)	1.73(0.62)	0.39(0.16)
	Scenario 3				
p_T	0.01	0.05	0.1	0.15	0.3
p_I	0.2	0.55	0.56	0.57	0.58
p_E	0.5	0.6	0.55	0.45	0.25
Desirability	50	**90**	55	55	45
ITIT	15.41(8.85)	**68.0(17.37)**	14.45(3.20)	2.06(0.50)	0.08(0.07)
STEIN	17.31(8.6)	**43.56(11.3)**	27.78(5.9)	10.95(3.0)	0.40(1.1)
BOIN	0.03(3.14)	**0.35(3.70)**	2.96(4.85)	26.60(7.33)	70.06(10.98)
3 + 3	2.88(3.71)	**9.60(5.17)**	18.07(6.20)	39.28(7.97)	30.01(6.90)
	Scenario 4				
p_T	0.15	0.2	0.33	0.38	0.43
p_I	0.2	0.55	0.56	0.57	0.58
p_E	0.2	0.6	0.62	0.66	0.68
Desirability	25	**90**	32	32	32
ITIT	14.52(8.36)	**78.75(19.44)**	5.48(1.88)	0.25(0.08)	0.01(0.00)
STEIN	15.83(7.8)	**66.36(15.9)**	13.97(5.1)	2.18(0.8)	0.18(0.1)
BOIN	7.56(7.25)	**33.08(9.85)**	35.22(8.12)	16.34(3.34)	6.57(1.16)
3 + 3	25.76(9.32)	**32.73(9.08)**	14.65(4.52)	5.36(1.63)	1.54(0.54)
	Scenario 5				
p_T	0.05	0.1	0.15	0.25	0.4
p_I	0.2	0.25	0.75	0.38	0.35
p_E	0.1	0.3	0.6	0.55	0.4
Desirability	25	25	**90**	50	11
ITIT	5.38(4.54)	6.47(7.08)	**86.37(17.86)**	1.69(0.43)	0.06(0.08)
STEIN	1.51(4.2)	10.53(6.1)	**65.41(13.4)**	21.35(5.2)	1.06(1.2)
BOIN	0.26(3.72)	2.35(4.91)	**16.32(6.95)**	48.06(8.53)	32.99(5.89)
3 + 3	9.26(5.38)	18.08(6.82)	**31.32(7.86)**	29.15(6.29)	9.51(2.98)
	Scenario 6				
p_T	0.05	0.1	0.15	0.32	0.5
p_I	0.12	0.2	0.8	0.81	0.83
p_E	0.2	0.4	0.45	0.47	0.5
Desirability	25	25	**55**	19	19
ITIT	4.76(4.24)	18.78(7.53)	**75.25(17.77)**	1.17(0.45)	0.01(0.00)
STEIN	5.08(4.8)	31.53(8.7)	**46.70(9.5)**	15.92(5.8)	0.74(1.3)
BOIN	0.26(3.72)	2.38(4.92)	**28.01(8.39)**	56.14(9.37)	13.19(3.60)
3 + 3	9.59(5.44)	17.74(6.74)	**42.68(9.52)**	24.20(5.91)	3.21(1.75)
	Scenario 7				
p_T	0.05	0.1	0.15	0.2	0.27
p_I	0.1	0.12	0.2	0.8	0.3
p_E	0.05	0.1	0.15	0.65	0.45
Desirability	25	25	25	**90**	50
ITIT	9.52(3.86)	7.52(5.17)	7.57(7.02)	**74.92(13.76)**	0.44(0.18)
STEIN	4.34(4.7)	10.99(5.8)	13.90(5.7)	**63.60(11.5)**	5.93(2.2)
BOIN	0.26(3.72)	2.36(4.91)	10.59(6.17)	**26.45(6.64)**	60.32(8.56)
3 + 3	9.38(5.40)	17.46(6.73)	21.99(6.60)	**23.98(5.54)**	24.46(5.05)

(*continued on next page*)

Table 2 (*continued*).

	Dose level				
	1	2	3	4	5
	Scenario 8				
p_T	0.05	0.08	0.12	0.15	0.35
p_I	0.1	0.2	0.25	0.85	0.7
p_E	0.2	0.3	0.4	0.45	0.4
Desirability	25	25	25	**55**	16
ITIT	6.18(4.09)	13.09(5.83)	19.76(7.92)	**60.65(12.01)**	0.29(0.15)
STEIN	4.98(4.8)	16.09(6.1)	32.89(7.7)	**40.94(7.7)**	5.07(3.7)
BOIN	0.09(3.63)	0.98(4.28)	4.52(5.16)	**35.21(7.56)**	59.18(9.37)
3 + 3	6.45(4.78)	12.00(5.64)	16.17(5.73)	**42.42(7.87)**	20.22(5.29)
	Scenario 9				
p_T	0.05	0.05	0.05	0.1	0.1
p_I	0.06	0.07	0.08	0.1	0.1
p_E	0.01	0.2	0.3	0.35	0.8
Desirability	10	10	10	25	**80**
ITIT	1.31(3.60)	1.94(3.97)	4.17(4.61)	4.73(5.56)	**87.82(12.25)**
STEIN	0.02(3.7)	2.83(4.4)	7.99(5.1)	7.66(5.3)	**81.31(11.5)**
BOIN	0.01(3.56)	0.09(3.56)	0.58(3.72)	2.93(4.39)	**96.37(14.77)**
3 + 3	2.48(3.93)	2.63(3.78)	8.80(4.52)	9.72(4.51)	**73.75(12.61)**
	Scenario 10				
p_T	0.1	0.1	0.1	0.1	0.1
p_I	0.05	0.06	0.08	0.1	0.5
p_E	0.18	0.2	0.23	0.25	0.7
Desirability	10	10	10	25	**100**
ITIT	7.53(4.84)	4.26(4.81)	3.95(4.79)	3.62(4.61)	**80.39(10.89)**
STEIN	7.40(6.0)	8.61(5.7)	8.01(4.9)	7.87(4.1)	**67.51(9.3)**
BOIN	0.63(4.62)	1.58(4.46)	3.24(4.36)	4.97(4.12)	**89.27(12.37)**
3 + 3	8.73(5.58)	7.66(4.72)	6.97(4.01)	7.50(3.55)	**59.40(9.71)**

sample size. I also replaced $\phi_I = 0.5$ and $\phi_E = 0.7$ with $\phi_I = \phi_E = 0.3$ so that I could see the performance when the target rates for immune response and overall response were low. Web Table 6 showed that the ITIT design was rather robust to the target rates.

For simplicity to specify the interval boundaries, I assumed independence among outcomes to derive optimal boundaries. Thus, I evaluated the robustness of the proposed method to dependence between multiple clinical outcomes in Web Appendix B. I generated correlated outcomes Y_T, $\tilde{Y}_I$, and Y_E which showed weak (between 0.3 and 0.5), moderate (between 0.5 and 0.7), and strong (>0.7) correlations. The proposed methods with the optimal boundaries (1), (2), and (3) were applied to the simulated data for both independent case and dependent cases. Web Table 8 showed that the proposed method assuming independence among endpoints worked well for the correlated outcomes. When toxicity and immune response were more correlated, immune response probability was increased. The selection percentage at OBD was 1%–5% smaller than the independent case, but the number of patients treated at OBD was similar regardless of the correlation between toxicity and immune response. When the immune response and tumor response were more correlated, the selection percentage at OBD was 0.5%–7% larger than in the independent case and one more patient was treated at OBD than in the independent case. When the toxicity and tumor response were more correlated, the selection percentage at OBD was 0.5%–7% larger than in the independent case but the number of patients treated at OBD was similar regardless of the correlation between toxicity and tumor response. I also compared the performance of ITIT with other designs. ITIT still worked better than other designs when multiple outcomes were correlated. It suggested that the optimal boundaries assuming independence worked well for the performance of the design.

The proposed allocation rule had one region for dose escalation, which seemed conservative. Thus, I considered several modified dose allocation rules which accelerated dose exploration and investigated the performance of ITIT. The results were presented in Web Appendix C. The proposed dose allocation rule mostly worked better than the less conservative rules under ITIT. When the OBD was the highest

dose (i.e., scenarios 9–10), the less conservative rules sped to escalate the dose and 1–3 patients were able to get benefit more than the proposed ITIT rule. However, in scenarios 1–8, the proposed allocation rule treated 7.8 patients more at OBD than the less conservative rules (minimum 3.82 and maximum 14.48). In all scenarios 1–10, the proposed ITIT showed an average of 14.20% more to select OBD than the less conservative rules (minimum 3.07 and maximum 25.2).

As noted, the simulation studies were performed with binary outcomes, which were generated from the Bernoulli distribution with the response probabilities (p_T, p_I, p_E) given at each dose. These could be done by the probability models, i.e., by regarding the probabilities as the function of dose and clinical outcomes. The model-based designs provided the probability models to generate the clinical outcomes of immunotherapy [5,7]. Web Appendix D provided the description of the probability models of Guo et al. [7] and Liu et al. [5] and the simulation results comparing ITIT with STEIN, BOIN, and 3+3 design.

ITIT could be extended to handle late-onset outcomes by using the imputation method [12] proposes. I investigated the performance of ITIT using the imputation method and provided results in Web Appendix E. ITIT with the delayed toxicity outcome still worked well and showed comparable accuracy to identify the OBD. ITIT with the delayed toxicity outcome also performed well compared STEIN, BOIN, and 3+3 design in most scenarios when all outcomes were available.

4. Discussion

I have proposed an interval design based on toxicity, immune response, and tumor response for dose-finding trials. In this proposed design, multiple clinical outcomes for immunotherapy trials are incorporated to allocate the next cohort dose and to find the recommended dose (i.e., OBD). By effectively using clinical outcomes, the design addresses limitations and challenges in the traditional paradigm for immunotherapy. Moreover, the newly proposed algorithm to allocate the dose is more appropriate and does not require any complicated statistical analysis. Desirability-based methods are also proposed for immunotherapy to determine OBD based on all outcomes. In the simulation study, I evaluated the operating characteristic of the proposed design and compared with other dose-finding designs (i.e., the STEIN, BOIN, and 3+3 design). Compared to the STEIN design using toxicity and tumor response, ITIT showed efficiency gain from using immune information. Compared to the BOIN and 3+3 design using only toxicity, ITIT performed better in most scenarios.

A tailored definition of OBD is of interest to practice. Based on the study objective, the proposed methods can be applied appropriately for less than three outcomes (e.g., toxicity and tumor response) or more than three outcomes. In addition, when the study team or sponsor wants to focus on specific events (e.g., some combinations of toxicity, immune response, or tumor response) and measure the associated event outcome, the desirability-based methods are flexible to use for the situation to identify OBD based on the clinical benefits and harms.

The clinical outcomes can be delayed to observe in immunotherapy trials and I investigated the performance of ITIT with the delayed toxicity outcome through simulations. I used the imputation method for delayed outcomes to make the decision of the dose allocation. Alternatively, if surrogate endpoints exist for the endpoints of interest and they are measured during the trial, surrogate endpoints can help as ancillary outcomes to assist with dose allocation and continual of the trial. Fleming [18] and Roep and Peakman [19] say that the surrogate endpoints take into account the delayed effect for the particular dose and can increase the chance of characterizing the immonotherapeutic profiles. A possible approach is to make the decision in a conventional way based on surrogate endpoints assuming that the outcomes are available. However, it requires caution to adjust the bias due to replacing the primary endpoint with the surrogate endpoint. It is not well investigated and worthy of developing the methods for the late onset outcomes.

In addition, as a future research, the proposed design can be extended to handle drug-combination trials. ITIT for drug-combination trials will use the same escalation/de-escalation rule as the single-agent trials proposed in this paper so that it maintains the simplicity to implement the design. However, since it is a two-dimensional dose-finding study, it requires to define admissible dose escalation/de-escalation sets. Park and Liu [20] considers two types for dose movements of the drug-combination trials: ND-design prohibits diagonal dose movements and D-design allows diagonal dose movements. Following either ND-design or D-design for the dose movements, dose assignment algorithms are established to maintain the feature of interval designs and have higher accuracy to identify the OBD.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The author thanks the editor and two referees for helpful suggestions. Dr. Park is partially supported by University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education.

Appendix A. Identification of interval boundaries of ITIT

It is critically important to identify the interval boundaries $\lambda_1, \lambda_2, \eta$, and δ, because the operating characteristics of the proposed design are determined by the interval boundaries. I adopt the approach in [9] to identify boundaries for the interval design. Based on the observed data and thresholds, I classify doses into (overly) safe, plausible (i.e., close to target toxicity rate), and overly toxic in terms of safety; I classify doses into immune activated and inactivated in terms of immune response; I classify doses into efficacious and inefficacious in terms of overall response. I want to minimize the misclassification rate in terms of safety, overall response, and immune response. For simplicity to derive the boundaries, I assume the independence among clinical outcomes.

For each $j = 1, \ldots, J$, let $p_{T,j}$, $p_{I,j}$, and $p_{E,j}$ denote the true probability for toxicity, immune response, and overall response, respectively, of dose level j. I first formulate three hypotheses at dose level j as follows:

$$H_{1j} : p_{T,j} = \phi_{T,1} \quad \text{and} \quad H_{2j} : p_{T,j} = \phi_T \quad \text{and}$$
$$H_{3j} : p_{T,j} = \phi_{T,2}.$$

The values of $\phi_{T,1}$ and $\phi_{T,2}$ partition the support of $p_{T,j}$ into three subintervals classifying the dose (1) dose is (overly) safe if $p_{T,j} \le \phi_{T,1}$, (2) dose has a close toxicity probability to target toxicity rate if $\phi_{T,1} < p_{T,j} < \phi_{T,2}$ (3) dose is overly toxic if $p_{T,j} \ge \phi_{T,2}$. Let λ_1 and λ_2 be the lower and upper thresholds for the observed probability $\hat{p}_{T,j}$ to classify the dose into (overly) safe, plausible, and overly toxic categories. Then, the posterior probability of misclassification of the current dose in terms of toxicity is given by

$$\epsilon_1 = \Pr(H_{1j}) \Pr(\hat{p}_{T,j} > \lambda_1 | H_{1j}) + \Pr(H_{2j}) \Pr(\hat{p}_{T,j} \le \lambda_1 \text{ or } \hat{p}_{T,j} \ge \lambda_2 | H_{2j}) + \Pr(H_{3j}) \Pr(\hat{p}_{T,j} < \lambda_2 | H_{3j}).$$

Assume that the prior probabilities for hypotheses are the same, i.e., $\Pr(H_{1j}) = \Pr(H_{2j}) = \Pr(H_{3j}) = 1/3$. Then, the optimal values of λ_1 and λ_2 are identified as

$$\lambda_1 = \frac{\log\{(1-\phi_{T,1})/(1-\phi_T)\}}{\log[\phi_T(1-\phi_{T,1})/\{\phi_{T,1}(1-\phi_T)\}]} \quad \text{and}$$
$$\lambda_2 = \frac{\log\{(1-\phi_T)/(1-\phi_{T,2})\}}{\log[\phi_{T,2}(1-\phi_T)/\{\phi_T(1-\phi_{T,2})\}]} \tag{1}$$

by minimizing the misclassification probability ϵ_1 [9,11].

I now formulate two hypotheses $H_{4j} : p_{E,j} = \phi_E$ and $H_{5j} : p_{E,j} = \phi_{E,1}$. The cutoff $\phi_{E,1}$ partitions a unit interval for $p_{E,j}$ into two subintervals, classifying the dose into efficacious and inefficacious doses if $p_{E,j} > \phi_{E,1}$ and $p_{E,j} \le \phi_{E,1}$, respectively.

Based on the observed data, the dose is classified in terms of efficacy by comparing $\hat{p}_{E,j}$ with threshold δ. Thus, the posterior probability of misclassification of the dose level j in regard to efficacy is given by

$$\epsilon_2 = \Pr(H_{4j}) \Pr(\hat{p}_{E,j} \le \delta | H_{4j}) + \Pr(H_{5j}) \Pr(\hat{p}_{E,j} > \delta | H_{5j}).$$

When I specify an equal prior probability, i.e., $\Pr(H_{4j}) = \Pr(H_{5j}) = 1/2$, the boundary

$$\delta = \frac{\log\{(1-\phi_{E,1})/(1-\phi_E)\}}{\log[\phi_E(1-\phi_{E,1})/\{\phi_{E,1}(1-\phi_E)\}]} \tag{2}$$

minimizes the misspecification error ϵ_2 due to overall response. The derivation of the optimal values of δ is similar to the derivation of λ_1.

Similarly for the overall response, I consider the following two hypotheses: $H_{6j} : p_{I,j} = \phi_I$ and $H_{7j} : p_{I,j} = \phi_{I,1}$. The value of $\phi_{I,1}$ partitions the unit interval for $p_{I,j}$ into two subregions $[0, \phi_{I,1}]$ and $(\phi_{I,1}, 1]$ to denote clinically uninteresting immune response and clinically desired immune response, respectively. Let η be the threshold for immune outcome based on the observed data. Then, the optimal value of η is obtained by minimizing the posterior probability of misclassification of the current dose level j in regard to immune response given by

$$\epsilon_3 = \Pr(H_{6j}) \Pr(\hat{p}_{I,j} \le \eta | H_{6j}) + \Pr(H_{7j}) \Pr(\hat{p}_{I,j} > \eta | H_{7j}).$$

For simplicity to find the minimizer of ϵ_3 with respect to η, an equal prior probability (i.e., $\Pr(H_{6j}) = \Pr(H_{7j}) = 1/2$) is considered, and the optimal boundary η is identified as

$$\eta = \frac{\log\{(1-\phi_{I,1})/(1-\phi_I)\}}{\log[\phi_I(1-\phi_{I,1})/\{\phi_{I,1}(1-\phi_I)\}]}. \tag{3}$$

The boundaries (1)–(3) are derived by minimizing each error rate ϵ_1, ϵ_2, or ϵ_3. Under the independence assumption among clinical outcomes, overall decision error rate is the sum of ϵ_1, ϵ_2, and ϵ_3. Therefore, the proposed ITIT design is optimal in the sense that it yields the minimum overall decision error rate. Cunanan and Koopmeiners [21] shows the independent modeling among toxicity and overall response performs as good as the joint model in small sample trials. Moreover, I checked in simulation study ignoring the correlation among outcomes has little impact on the performance of the design, and the sensitivity analysis showed the robustness of optimal boundaries to dependence between clinical outcomes (See Web Appendix B).

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.conctc.2022.101005. Web Appendices A–E are available online with this paper. The programming code is available at the author's personal webpage.

References

[1] W. Alexander, The checkpoint immunotherapy revolution: what started as a trickle has become a flood, despite some daunting adverse effects; new drugs, indications, and combinations continue to emerge, Pharm. Ther. 41 (3) (2016) 185.

[2] J. Arrowsmith, Trial watch: phase iii and submission failures: 2007–2010, 2011.

[3] B. Seruga, A. Ocana, E. Amir, I.F. Tannock, Failures in phase iii: causes and consequences, 2015.

[4] FDA, 22 Case studies where phase 2 and phase 3 trials had divergent results, 2017.

[5] S. Liu, B. Guo, Y. Yuan, A Bayesian phase I/II trial design for immunotherapy, J. Amer. Statist. Assoc. 113 (523) (2018) 1016–1027.

[6] B. Guo, D. Li, Y. Yuan, SPIRIT: A seamless phase i/II randomized design for immunotherapy trials, Pharm. Statist. 17 (5) (2018) 527–540.

[7] B. Guo, Y. Park, S. Liu, A utility-based Bayesian phase I–II design for immunotherapy trials with progression-free survival end point, J. R. Stat. Soc. Ser. C. Appl. Stat. 68 (2) (2019) 411–425.

[8] K. Takeda, M. Taguri, S. Morita, BOIN-ET: Bayesian optimal interval design for dose finding based on both efficacy and toxicity outcomes, Pharm. Statist. (2018).

[9] R. Lin, G. Yin, STEIN: A simple toxicity and efficacy interval design for seamless phase I/II clinical trials, Stat. Med. 36 (26) (2017) 4106–4120.

[10] L. Seymour, J. Bogaerts, A. Perrone, R. Ford, L.H. Schwartz, S. Mandrekar, N.U. Lin, S. Litière, J. Dancey, A. Chen, et al., IRECIST: guidelines for response criteria for use in trials testing immunotherapeutics, Lancet Oncol. 18 (3) (2017) e143–e152.

[11] S. Liu, Y. Yuan, Bayesian optimal interval designs for phase I clinical trials, J. R. Stat. Soc. Ser. C. Appl. Stat. 64 (3) (2015) 507–523.

[12] Y. Yuan, R. Lin, D. Li, L. Nie, K.E. Warren, Time-to-event Bayesian optimal interval design to accelerate phase i trials, Clin. Cancer Res. 24 (20) (2018) 4921–4930.

[13] Y. Zang, J.J. Lee, Y. Yuan, Adaptive designs for identifying optimal biological dose for molecularly targeted agents, Clin. Trials 11 (3) (2014) 319–327.

[14] M.-K. Riviere, Y. Yuan, J.-H. Jourdan, F. Dubois, S. Zohar, Phase I/II dose-finding design for molecularly targeted agent: Plateau determination using adaptive randomization, Stat. Methods Med. Res. 27 (2) (2018) 466–479.

[15] J. O'Quigley, M. Pepe, L. Fisher, Continual reassessment method: a practical design for phase 1 clinical trials in cancer, Biometrics (1990) 33–48.

[16] Y. Yuan, K.R. Hess, S.G. Hilsenbeck, M.R. Gilbert, Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials, Clin. Cancer Res. 22 (17) (2016) 4291–4301.

[17] H. Zhou, Y. Yuan, L. Nie, Accuracy, safety, and reliability of novel phase I trial designs, Clin. Cancer Res. 24 (18) (2018) 4357–4364.

[18] T.R. Fleming, Surrogate endpoints and FDA's accelerated approval process, Health Aff. 24 (1) (2005) 67–78.

[19] B.O. Roep, M. Peakman, Surrogate end points in the design of immunotherapy trials: emerging lessons from type 1 diabetes, Nat. Rev. Immunol. 10 (2) (2010) 145–152.

[20] Y. Park, S. Liu, On the coherence of model-based dose-finding designs for drug combination trials, PLoS One 15 (11) (2020) e0242561.

[21] K. Cunanan, J.S. Koopmeiners, Evaluating the performance of copula models in phase I-II clinical trials under model misspecification, BMC Med. Res. Methodol. 14 (1) (2014) 51.