SPECIAL SERIES: STATISTICAL METHODS IN PRECESION ONCOLOGY

review articles

Model-Assisted Designs for Early-Phase Clinical Trials: Simplicity Meets Superiority

Ying Yuan, PhD[1]; J. Jack Lee, PhD[1]; and Susan G. Hilsenbeck, PhD[2]

abstract

Drug development enterprise is struggling because of prohibitively high costs and slow progress. There is urgent need for adoption of novel adaptive designs to improve the efficiency and success of clinical trials. A major barrier is that many conventional designs are inadequate for modern drug development, yet most novel adaptive designs are difficult to understand, require complicated statistical modeling, demand complex computation, and need expensive infrastructure for implementation. The objective of this article is to introduce and review a class of novel adaptive designs, known as model-assisted designs, to remove this barrier and increase the use of novel adaptive designs. Model-assisted designs enjoy superior performance comparable to more complicated, model-based adaptive designs, but their decision rule can be pretabulated and included in the protocol—thus implemented as simply as the conventional designs. We review state-of-the-art model-assisted designs for phase I clinical trials for single-agent, drug-combination and late-onset toxicity scenarios. We also briefly introduce model-assisted designs for phase II trials to handle binary, coprimary endpoints and delayed response. Freely available user-friendly software and trial examples (trialdesign.org) facilitate the adoption of model-assisted designs.

INTRODUCTION

Despite rapid advancements in the knowledge of biomedicine, current drug development is on an unsustainable path plagued by high costs, slow progress, and a high failure rate.[1,2] A recent survey reports that the estimated average out-of-pocket cost per approved new drug is 2.5 billion US dollars.[3] To address this pressing issue, one important approach embraced by the US Food and Drug Administration is to use novel adaptive designs.[4]

Numerous novel adaptive designs have been proposed to improve the efficiency and accuracy of phase I trials to find the maximum-tolerated dose (MTD) and of phase II trials to identify effective treatments.[5-7] Most of these novel designs, however, failed to be translated into clinical trials,[8,9] because they often are difficult to understand and require complicated statistical modeling, demanding computation, and expensive infrastructure for implementation. As a result, conventional designs (eg, the 3 + 3 design) are still dominantly used despite relatively poor performance. There is an urgent need to increase the adoption of novel adaptive designs.

The objective of this article is to introduce and review a class of novel phase I and II designs, known as model-assisted designs,[10-12] to overcome this quandary of simplicity versus performance. Model-assisted designs yield superior performance compared with the conventional algorithm-based designs and are comparable to more complicated (model-based) designs. With the model-assisted designs, the decision rule can be pretabulated and included in the protocol and, thus, implemented in as simple a way as the conventional designs. By increasing the awareness of model-assisted designs, we hope that more practitioners will apply the novel designs to improve the efficiency and success of early-phase trials.

PHASE I TRIAL DESIGNS

On the basis of their statistical foundation and implementation approach, phase I trial designs can be classified into three types: algorithm-based, model-based, and model-assisted designs.[10-12] Table 1 contrasts the characteristics of these three types of designs.

ALGORITHM-BASED DESIGNS

Algorithm-based design is a class of conventional design that uses a set of simple, prespecified rules to determine the dose escalation and de-escalation. Examples include the conventional 3 + 3 design[13] and its extensions, such as the accelerated titration design[14] and the rolling 6 design.[15] It has long been known that the 3 + 3 design has relatively poor operating characteristics[16,17]; for example, it has no specific target dose-limiting toxicity (DLT) rate (but has a range of DLT rates between 17% and 33%), has poor accuracy to identify the MTD, has poor precision to

ASSOCIATED CONTENT

Appendix

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on May 8, 2019 and published at ascopubs.org/journal/po on October 24, 2019: DOI https://doi.org/10.1200/PO.19.00032

CONTEXT

Key Objective

Is there any novel adaptive design that is simple to implement?

Knowledge Generated

This article introduces and reviews a class of novel phase I and II designs, known as model-assisted designs, to provide a state-of-the-art approach to overcome the quandary of simplicity versus performance that hinders the adoption of novel adaptive designs. Model-assisted designs yield superior performance compared with more complicated model-based designs, but the decision rule can be pretabulated and included in the protocol and, thus, implemented in as simple a way as the conventional designs.

Relevance

Model-assisted designs are easy to implement and have great potential to improve the efficiency and success rate of early-phase trials.

estimate the DLT rate, and has a greater tendency to underdose patients (ie, it treats patients at the doses lower than the MTD). However, because it is simple and easy to implement, the 3 + 3 design is by far the most commonly used phase I design in practice.

MODEL-BASED DESIGNS

Model-based design is a class of novel adaptive designs that uses a statistical model (eg, a logistic model) to describe the dose-toxicity curve and guide dose transition. Examples include the continuous reassessment method (CRM)[17] and its various extensions (eg, dose escalation with overdose control),[18] Bayesian logistic regression model,[19] and Bayesian model averaging CRM.[20] As information accrues during the trial, the CRM continuously updates the estimate of the model after each cohort and then uses the updated estimate to determine the dose for the next cohort. Numerous studies have shown that the CRM significantly outperforms the 3 + 3 design,[16,21] with higher accuracy to identify and allocate more patients to the MTD as well as the ability to target any prespecified DLT rates. Despite decades of advocacy by statisticians, the use of the CRM, however, is still limited because of statistical and computational complexity of the design.[8,9] For appropriate use, the CRM requires specialized expertise to choose and calibrate the dose-toxicity model and to re-estimate the model at each decision of dose escalation/de-escalation. It remains a challenge to communicate to clinicians how the design works, which leads them to perceive dose allocations as coming from a black box.

MODEL-ASSISTED DESIGNS

Model-assisted designs were developed to combine the advantages of algorithm-based designs and model-based designs.[10,12,22] Similar to the model-based design, the model-assisted design uses a statistical model (eg, the binomial model) to derive the design for efficient decision making; however, like the algorithm-based design, its dose escalation and de-escalation rule can be predetermined before the onset of the trial and, thus, can be implemented in as simple a way as the algorithm-based designs. Examples of model-assisted designs include the modified toxicity probability interval (mTPI) design[23] and its variation, mTPI-2[24]; Bayesian optimal interval (BOIN) design[25,26]; and keyboard design[22] (Fig 1). Zhou et al[11,12] and Ruppert and Shoben[27] conducted comprehensive numeric studies to compare the model-assisted designs with the 3 + 3 design and several model-based designs (eg, CRM and escalation with overdose control).[12] The results showed that the model-assisted designs substantially outperformed the 3 + 3 design and yielded a performance comparable to model-based designs on several metrics, including the accuracy of identification of the MTD and allocation of patients to the MTD and the risk of overdosing patients (ie, treatment of a patient at a dose greater than the MTD). Among the model-assisted designs, BOIN stands out; it outperforms the mTPI with higher accuracy identifying the MTD and a lower risk of overdosing patients, and it is simpler and more transparent than the mTPI-2 and keyboard designs.[12] BOIN is also more versatile; it can handle drug-combination trials,[28] late-onset toxicity,[29] low-grade toxicities,[30] and toxicity and efficacy jointly.[31,32] Therefore, in what follows, after a brief review of the mTPI/mTPI-2 and keyboard designs, we use BOIN as an example to illustrate the features and advantages of model-assisted designs.

The mTPI design starts with a definition of three toxicity probability intervals: underdosing, proper dosing, and overdosing intervals—for example, (0, 0.2), (0.2, 0.4), and (0.4, 1), respectively. Given the data observed at the current dose, mTPI makes the decision of dose escalation and de-escalation on the basis of the unit probability mass (UPM) of the three intervals. Let p denote the true DLT probability of the current dose. The UPM of an interval is defined as the posterior probability that p is within the interval divided by the length of the interval, calculated according to a statistical model known as the beta-binomial model. If the UPM associated with the underdosing (or overdosing) interval is the largest among the three UPMs, the design escalates (or de-escalates) the dose; otherwise,

TABLE 1. Comparison of Design Characteristics Among Algorithm-Based, Model-Based, and Model-Assisted Phase I Designs

Design Characteristic	Algorithm Based	Model Assisted	Model Based
Transparency and simplicity			
Dose escalation/de-escalation rule can be predetermined and included in the protocol	Yes	Yes	No
Avoids computation-intensive, repeated estimation of the dose-toxicity curve model to make interim decisions	Yes	Yes	No
Flexibility			
Targets any prespecified DLT rate	No	Yes	Yes
Allows decision making when the cohort size deviates from the planned size	No	Yes	Yes
No. of patients treated at the MTD can be > 6	No	Yes	Yes
Sample size can be calibrated to ensure good operating characteristics	No	Yes	Yes
Performance			
Identifies the MTD accurately	No	Yes	Yes
Allocates a high percentage of patients to the MTD	No	Yes	Yes
Provides good overdose control	Yes	Yes	Yes

Abbreviations: DLT, dose-limiting toxicity; MTD, maximum-tolerated dose.

the dose stays at the current dose. One deficiency of mTPI is that it overly downweighs the overdosing probability, because the overdosing interval is typically wider than the proper dosing interval, which leads to a high risk of overdose of patients (ie, treatment of a high percentage of patients at doses greater than the MTD).[22]

The keyboard design addresses the overdosing issue of mTPI by defining a series of equal-width dosing intervals (or keys) to represent the potential locations of p.[22] The proper dosing interval is called the target key. The design makes the decision of dose escalation and de-escalation by examining the relative position between the target key and the strongest key, in which the strongest key is defined as the interval that the true DLT is most likely located. The strongest key is identified using the beta-binomial model. If the strongest key is on the left (or right) side of the target key, the observed data suggest that the current dose is most likely underdosing (or overdosing); thus, the design escalates (or de-escalate) the dose; otherwise, the dose stays at the current dose. The keyboard design outperforms the mTPI with substantially lower risk of overdosing patients and better accuracy to identify the MTD.[11,22] The variation of the mTPI (ie, mTPI-2[24]) adopts the same dose escalation/de-escalation rule as the keyboard design but is less transparent. The mTPI-2 relies on complicated procedures, such as Occam's razor and model selection.

Compared with the mTPI/mTPI-2 and keyboard designs, the BOIN design is more straightforward and transparent (Fig 1). Let $\hat{p}$ denote the observed DLT rate at the current dose, defined as the number of patients experiencing DLT at the current dose divided by the total number of DLT-evaluable patients treated at the current dose.

The BOIN design makes dose escalation/de-escalation recommendations simply by comparing $\hat{p}$ with prespecified dose escalation (λ_e) and de-escalation (λ_d) boundaries, as illustrated in Figure 2 and described as follows:

1. Treat the first cohort of patients at the lowest dose, or the clinician-specified starting dose.
2. Assign a dose to the next cohort of patients:
   - If $\hat{p} \le \lambda_e$, escalate the dose to the next higher level.
   - If $\hat{p} \ge \lambda_d$, de-escalate the dose to the next lower level.
   - Otherwise, stay at the current dose.
3. Repeat step 2 until the prespecified maximum sample size is reached or the number of patients treated on a single level reaches a certain number (eg, 12). At that point, select the MTD as the dose at which the DLT estimate is closest to the target.[25,26]

Figure 2 provides the default, optimal dose escalation and de-escalation boundaries (λ_e, λ_d) for commonly used target DLT rates, φ. These boundaries minimize the incorrect decisions of escalating/de-escalating the dose when it actually is greater/lower than the MTD according to a binomial model.[26] Given that $\varphi = 0.3$, the escalation and de-escalation boundaries are $\lambda_e = 0.236$ and $\lambda_d = 0.358$, respectively. That is, at the current dose, if the observed DLT rate is less than 0.236 (eg, 0/3 DLT), we escalate the dose; if the observed DLT rate is greater than 0.358 (eg, 2/3 DLTs), we de-escalate the dose; otherwise, the dose stays at the same level (eg, 1/3 DLT). For patient safety, BOIN imposes an overdose control rule: If the observed data indicate that there is more than a 95% chance that the current dose is higher than φ and at least three patients have been treated, the current and higher doses are eliminated from the trial. The trial is terminated if the lowest dose is eliminated.

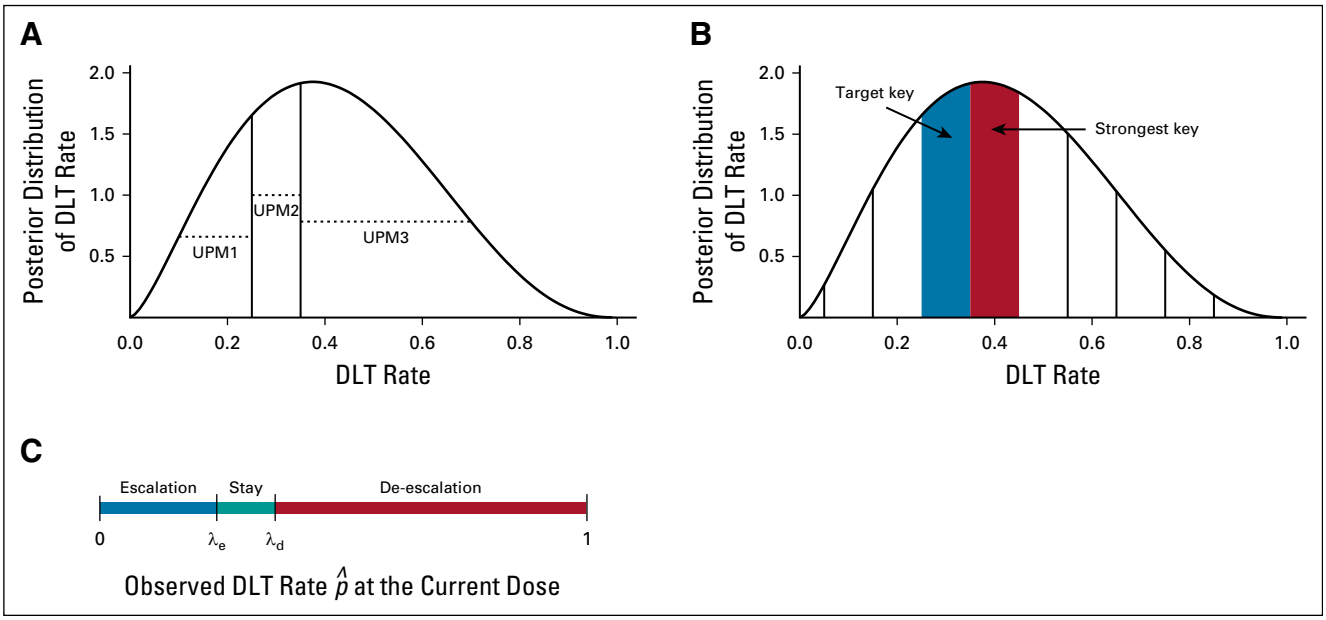


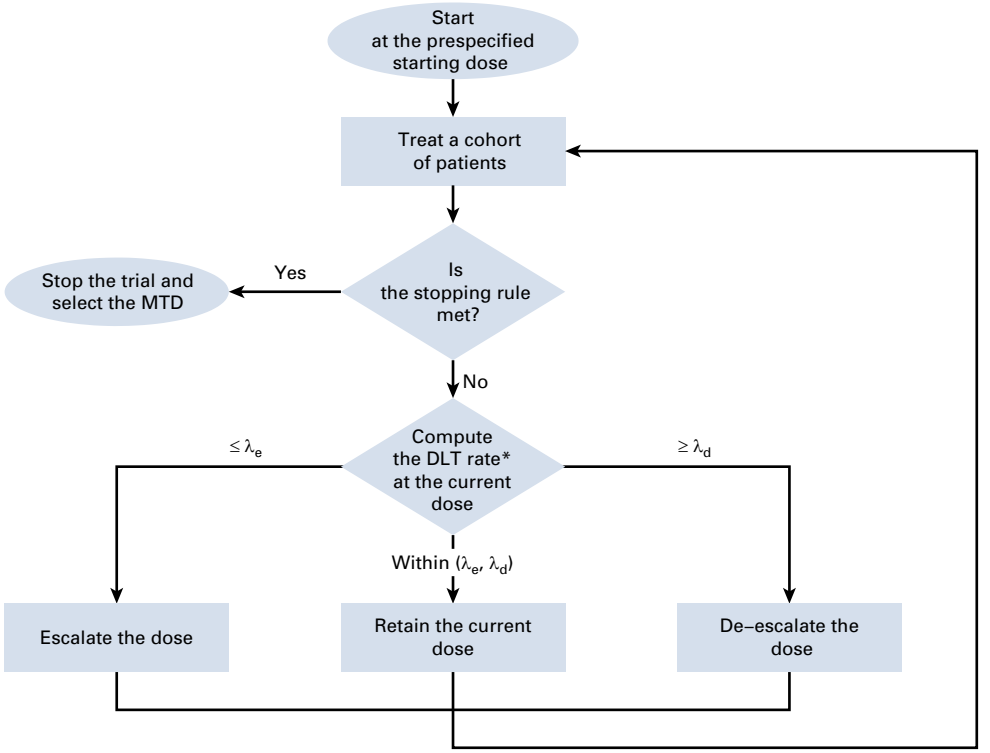
FIG 1. Illustration of the (A) modified toxicity probability interval, (B) keyboard, and (C) Bayesian optimal interval designs. The curves in (A) and (B) are the posterior distributions of the dose-limiting toxicity (DLT) probability at the current dose. To determine the next dose, the modified toxicity probability interval design calculates and compares the values of the three unit probability masses (UPMs), whereas the keyboard design compares the location of the strongest key with respect to the target key. Bayesian optimal interval compares the observed DLT rate at the current dose with the optimized dose escalation boundary, λ_e, and de-escalation boundary, λ_d.

The simple and intuitive structure of BOIN gives it several unique advantages. Because BOIN guarantees de-escalation of the dose when the observed toxicity rate $\hat{p}$ is higher than the de-escalation boundary λ_d, it is particularly easy for clinicians and regulatory agents to assess the safety of a trial with BOIN. For example, given a target DLT rate of $\varphi = 0.25$, we know a priori that a phase I trial using BOIN guarantees de-escalation of the dose if the observed DLT rate is higher than 0.298. Accordingly, BOIN also allows users to easily calibrate the design to satisfy specific safety requirements mandated by regulatory agents by choosing an appropriate target DLT rate. Suppose that, for a phase I trial with a new compound, the regulatory agency mandates that the dose must be de-escalated if the observed toxicity rate is higher than 0.25. We can easily fulfill that requirement by setting the target DLT rate at $\varphi = 0.21$, such that BOIN guarantees de-escalation of the dose if the observed toxicity rate of $\hat{p} \geq 0.25$. Such flexibility and transparency are the important advantages of BOIN compared with the other model-assisted designs, such as mTPI/mTPI-2 and keyboard designs.

Because model-assisted designs are built upon rigorous statistical theory, like model-based designs, they also enjoy the same flexibility as the model-based designs. For example, BOIN can target any prespecified DLT rate tailored to the trial. For heavily treated patients with recurrent cancer, a target DLT rate of greater than 0.3 may be an acceptable tradeoff to achieve greater treatment efficacy; conversely, for newly diagnosed patients with cancer, a lower target DLT rate (eg, 0.15 or 0.2) may be more appropriate. Similarly, we may tolerate higher rates of reversible DLTs, but drugs with serious irreversible toxicities may mandate lower target rates. In contrast, the 3 + 3 design has no specific target DLT rate, but we must find a dose with the DLT rate between 17% and 33%. In addition, unlike the 3 + 3 design, for which the dose escalation and de-escalation decisions can be made only when we have three or six evaluable patients, BOIN allows decision making with incomplete cohorts in the face of dropouts as a result of DLT inevaluability. This is because BOIN makes decisions on the basis of the observed DLT rate at the current dose, which can be calculated when given any number of observations.

TRIAL EXAMPLES

BOIN design has been used in variety of oncology trials including those for pediatric tumors[33,34] (ClinicalTrials.gov identifier: NCT02354547), adult tumors (ClinicalTrials.gov identifiers: NCT03577704, NCT0302316, NCT02942264, NCT03318900, NCT03600155, NCT0205075, NCT03740256, NCT03330028, NCT03114462, NCT03036904, NCT02705196, NCT02942095, NCT03740256, NCT03784677, and NCT03760081), solid tumors (eg, breast [ClinicalTrials.gov identifier: NCT0302316], brain [ClinicalTrials.gov identifier: NCT02942264], ovarian [ClinicalTrials.gov identifier: NCT03318900], stomach [ClinicalTrials.gov identifier: NCT03330028], neck [ClinicalTrials.gov identifier: NCT03114462], lung [ClinicalTrials.gov identifier: NCT02942095), bladder [ClinicalTrials.gov identifier: NCT03740256], prostate [ClinicalTrials.gov



	Target toxicity rate ϕ					
Boundaries	0.15	0.20	0.25	0.30	0.35	0.40
λ_e (escalation)	0.118	0.157	0.197	0.236	0.276	0.316
λ_d (de-escalation)	0.179	0.238	0.298	0.358	0.419	0.479

FIG 2. The flowchart of the Bayesian optimal interval design in which λ_e and λ_d are the dose escalation boundary and the de-escalation boundary, respectively. (*) DLT rate = total number of patients who experienced DLT at the current dose divided by the total number of patients treated at the current dose. MTD, maximum-tolerated dose.

identifier: NCT03784677], and germ cell [ClinicalTrials.gov identifier: NCT03760081]), and liquid tumors (eg, leukemia [ClinicalTrials.gov identifier: NCT03600155], and lymphoma NCT03114462]). BOIN design has been used for various treatment agents, including chemotherapy (ClinicalTrials.gov identifier:NCT02942264), radiotherapy (ClinicalTrials.gov identifier: NCT03114462), checkpoint inhibitor (ClinicalTrials.gov identifier: NCT03600155), monoclonal antibody (ClinicalTrials.gov identifier: NCT03577704), oncolytic virus (ClinicalTrials.gov identifier: NCT02705196), and T-cell immunotherapy (ClinicalTrials.gov identifier: NCT03318900). BOIN also has also been used in nononcology trials, such as stem cell therapy for stroke patients with stroke.[35]

We used an ongoing National Cancer Institute phase I to II trial (ClinicalTrials.gov identifier: NCT02942264) to illustrate the design. One of the trial's objectives of was to identify the MTD of TG02, a pyrimidine-based multikinase inhibitor combined with temozolomide in adult patients with recurrent anaplastic astrocytoma or glioblastoma. Because of lack of effective treatments for this patient population, the target DLT rate was set at a relatively high value of 0.35. Four doses (ie, 150, 200, 250, and 300 mg) of TG02 were investigated. The maximum sample size for dose finding was 24 patients, treated in cohorts of three patients each. According to the toxicity profile of TG02 in other patients with cancer, the principal investigator chose 200 mg of TG02 as the starting dose. According to the BOIN design, the dose will be escalated if the observed DLT rate at the current dose is lower than $\lambda_e = 0.276$, and it will be deescalated if the observed DLT rate is greater than $\lambda_d = 0.419$ (Fig 2). For the purpose of overdose control, if the observed data suggest that there is more than a 95% chance that a dose is greater than the MTD—that is, $\Pr(p > .35 \mid \text{data}) > 0.95$—that dose and

higher doses will be eliminated from additional examination. When the lowest dose is eliminated, the trial will be stopped for safety. The 3 + 3 design could not be used to find the MTD appropriately for this trial, because it deemed a DLT rate of 33% or greater to be unacceptably high.

LATE-ONSET TOXICITY

Late-onset toxicity is common in targeted therapies and immunotherapies.[36,37] It causes major logistic difficulty for aforementioned phase I designs, which require that a DLT must be observed soon enough to apply decision rules to choose doses for new patients. For example, if the DLT takes up to 8 weeks to evaluate and the accrual rate is one patient per week, on average, then five new patients could be accrued while investigators wait to evaluate the previous three patients' outcomes. The question is this: How can new patients receive timely treatment when the previous patients' outcomes are pending?

A few model-based designs have been proposed to address this logistic issue, including the time-to-event CRM (TITE-CRM)[38] and data-argumentation CRM (DA-CRM).[39] These designs support continuous accrual and yield superior performance in finding the MTD.[39,40] However, like the CRM, they are statistically and computationally complex and require repeated model fitting after each cohort, which limits their use.

Model-assisted designs, including time-to-event BOIN (TITE-BOIN)[29] and time-to-event keyboard (TITE-Keyboard) designs,[41] provide a well-performing and yet easy-to-implement approach to address late-onset toxicity. The TITE-BOIN works by predicting the unobserved, pending DLT data according to a time-to-event model. Thereby, the BOIN dose escalation/de-escalation rule described previously can be applied in real time to choose a dose for new patients. The TITE-BOIN uses a model for prediction, but its decision rule can be pretabulated, similar to the algorithm-based rolling 6 design. Table 2 shows the TITE-BOIN decision rule with a cohort size of three patients, which can be generated easily using the software described in the Software section. During the trial, at the current dose, we counted the number of patients, the number of patients who experienced DLT, and the number of pending patients and their standard total follow-up times (STFT); we then used the table to make the dose escalation/de-escalation decision. The STFT was computed as the sum of the follow-up times for all pending patients at the current dose divided by the DLT assessment window. To illustrate the use of the TITE-BOIN decision table, suppose that three patients have been treated at the current dose: one had a DLT, one had no DLT, and one has DLT data pending. According to Table 2, if the STFT of the pending patient is greater than 0.88, we treat the next cohort at the same dose; otherwise, we de-escalate the dose. Suppose that the next cohort of three patients is treated at the same dose and that, among the six treated patients, one patient had a DLT, two had no DLT, and three have DLT data pending. Given that the DLT assessment window is 3 months and that, at the current dose, three pending patients have been observed for 1, 1.6, and 2.5 months, respectively, the STFT is (1 + 1.6 + 2.5)/3 = 1.7. To treat the next cohort, because the STFT of the three patients with data pending is less than 1.96, we keep the current dose. Numeric study shows that TITE-BOIN yields superior performance compared with model-based designs (eg, TITE-CRM), and outperforms the rolling 6 design with higher accuracy to identify the MTD and allocate more patients to the MTD.[29]

Another model-assisted design that is capable of handling late-onset toxicity is the TITE-Keyboard design.[41] Rather than predicting the pending DLT data, the TITE-Keyboard design takes a different statistical approach by discounting the observed data information, statistically known as the likelihood, to reflect that some observations are pending. Nevertheless, the TITE-Keyboard produces a decision table similar to Table 2 and yields performance that is comparable to that of the TITE-BOIN.

DRUG COMBINATION TRIALS

Designing combination trials is more challenging. Unlike single-agent trials with a string of ordered doses, combinations in the dose matrix are only partially ordered in toxicity, and multiple MTDs (ie, the MTD contour) may exist in the dose matrix (Appendix Fig A1). Numerous designs have been proposed to find an MTD or the MTD contour for combination trials.[28,42-52] Almost all are model-based designs using a strategy similar to that of a CRM: devise a model to describe the dose-toxicity surface and then, on the basis of accumulating data, continuously update the model estimate to select a dose for the new patient. Because of their statistical and computational complexity, despite good performance, these model-based designs are rarely used for conducting trials.

Model-assisted designs provide a simple and robust approach to phase I combination trials.[43] One example is the BOIN combination design,[28] which makes the decision of dose escalation/de-escalation according to the same rule as the single-agent BOIN design described in the Model-Assisted Designs section, and thereby inherits the single-agent design's simplicity and good performance. The only difference is that, in combination trials, when we decide to escalate or de-escalate the dose, there is more than one neighboring dose to which we can move. For example, when we escalate/de-escalate the dose, we can escalate/de-escalate either the dose of drug A or the dose of drug B. The BOIN combination design makes this choice according to how likely a dose combination is to be located within the acceptable region (λ_e, λ_d) given the observed data.[28] Simulation study shows that the BOIN combination design yields competitive performance comparable to more

TABLE 2. Dose Escalation And De-Escalation Rule for TITE-BOIN With a Target DLT Rate of 0.3 and a Cohort Size of Three Patients, With up to Nine Patients Treated at a Dose

			Decision		
No. Treated	**No. With DLTs**	**No. With Data Pending**	**Escalate**	**Stay**	**De-Escalate**
3	0	≤ 1	Yes		
3	0	≥ 2		Suspend accrual	
3	1	0		Yes	
3	1	1		STFT > 0.88	STFT ≤ 0.88
3	1	≥ 2		Suspend accrual	
3	2	≤ 1			Yes
3	3	0			Yes and eliminate*
6	0	≤ 3	Yes		
6	0	≥ 4		Suspend accrual	
6	1	≤ 1	Yes		
6	1	2	STFT ≥ 0.60	STFT < 0.60	
6	1	3	STFT ≥ 1.96	STFT < 1.96	
6	1	≥ 4		Suspend accrual	
6	2	0		Yes	
6	2	1		STFT > 0.73	STFT ≤ 0.73
6	2	2		STFT > 1.80	STFT ≤ 1.80
6	2	3		STFT > 2.87	STFT ≤ 2.87
6	2	≥ 4		Suspend accrual	
6	3	≤ 3			Yes
6	≥ 4	≤ 2			Yes and eliminate*
9	0	≥ 4	Yes		
9	0	≥ 5		Suspend accrual	
9	1	≤ 4	Yes		
9	1	≥ 5		Suspend accrual	
9	2	0	Yes		
9	2	1	STFT ≥ 0.59	STFT < 0.59	
9	2	2	STFT ≥ 1.65	STFT < 1.65	
9	2	3	STFT ≥ 2.71	STFT < 2.71	
9	2	4	STFT ≥ 3.77	STFT < 3.77	
9	2	≥ 5		Suspend accrual	
9	3	0		Yes	
9	3	1		STFT > 0.58	STFT ≤ 0.58
9	3	2		STFT > 1.65	STFT ≤ 1.65
9	3	3		STFT > 2.72	STFT ≤ 2.72
9	3	4		STFT > 3.79	STFT ≤ 3.79
9	3	≥ 5		Suspend accrual	
9	4	≤ 5			Yes
9	≥ 5	≤ 4			Yes and eliminate*

NOTE. No. treated is the total number of patients treated at the current dose level; No. with DLTs is the number of patients who experienced DLT at the current dose level; No. with data pending denotes that number of patients whose DLT data are pending at the current dose level.

Abbreviations: DLT, dose-limiting toxicity; STFT, standard total follow-up time for the patients with data pending, defined as the total follow-up time for the patients with data pending divided by the length of the DLT assessment window; TITE-BOIN, time-to-event Bayesian optimal interval.

*When a dose is eliminated, all higher doses should also be eliminated.

complicated, model-based designs.[28] Another model-assisted combination design is the keyboard combination design, which adopts similar dose escalation/de-escalation strategy as the BOIN combination design and yields competitive performance.[52] Model-assisted designs also were developed to find multiple MTDs (or the MTD contour) for combination trials (eg, the waterfall design).[51]

MODEL-ASSISTED PHASE II DESIGNS

The concept of model-assisted design is also applicable to phase II trials. For example, Simon's optimal (or minimax) two-stage design[53] could be classified as a model-assisted design in the sense that it is derived from a statistical model (ie, a binomial model for response), but its go/no-go decision rule can be predetermined when the design parameters are specified. During the trial conduct, users only must evaluate whether the number of responses passes the predetermined stopping boundaries to make go/no-go decisions. Simon's optimal two-stage designs are appropriate for the simple setting in which the end point is binary and quickly ascertainable, and these designs allow only one interim look.

Phase II trials sometimes are more complicated and have more than one end point. Table 3 provides four trial examples with different types of end points (eg, ordinal end point and coprimary end points).[54-56] In addition, multiple interim looks are useful to improve the flexibility and efficacy of the trial, especially in basket and platform trials.[57-59] The Bayesian optimal phase II (BOP2) design[60] provides a simple, flexible, and efficient model-assisted design to allow multiple interims and handle different types of phase II trials under a unified framework.

The key feature of BOP2 is that, although the end points in the four examples are clinically different, they all can be represented using a variable Y with K distinct categories, statistically known as a multinomial random variable. For instance, in example 1, Y has $K = 2$ categories (1 = response and 2 = no response); in example 3, Y has $K = 4$ categories (ie, 1 = [response, progression-free survival at 6 months {PFS6} > 0.2 met]; 2 = [response, PFS6 not met]; 3 = [no response, PFS6 met]; and 4 = [no response, PFS6 not met]). This unified end point Y is modeled using a Bayesian model, known as the Dirichlet-multinomial model. The BOP2 design allows any arbitrary number of interim looks. At each interim look, the go/no-go decision is made on the basis of the evaluation of Bayesian stopping criteria using posterior probabilities. As an example, consider a treatment that is deemed ineffective if the objective response rate (ORR) is $\leq \theta$, in which θ is a threshold prespecified by clinicians (eg, $\theta = 20\%$). At each interim, the go/no-go decision is made according to the following Bayesian stopping criteria: Stop the trial if $\Pr(\text{ORR} \leq \theta \mid \textit{Data}) > C_n$; otherwise continue. In the equation, the posterior probability $\Pr(\text{ORR} \leq \theta \mid \textit{Data})$ represents, given the interim data, how likely the true ORR is to be less than the threshold θ; and C_n is an adaptive probability cutoff that depends on the interim sample size n. Given the response rate deemed ineffective (ie, the null hypothesis) and the response rate deemed desirable (ie, the alternative hypothesis), the value of C_n is chosen such that the type I error rate is controlled at a prespecified level and the statistical power is maximized. (See Zhou et al[60] for details.) Similar Bayesian stopping criteria can be applied to other types of end points to determine whether the treatment is promising.[60]

As a model-assisted design, one important advantage of the BOP2 design is that its stopping boundary can be enumerated and included in the trial protocol before the onset of the trial. Table 4 presents the corresponding stopping boundaries for each trial example. When they conduct the trial, clinicians simply count the number of relevant events and make the go/no-go decision according to whether that count exceeds the boundary or not. If the end point requires a long time to be scored, clinicians may have to suspend the accrual and wait for the interim data mature to make interim decisions. This is undesirable and prolongs the trial duration. The time-to-event BOP2 (TOP) design was developed to address this issue and allow real-time interim

TABLE 3. Four Examples Considered by the BOP2 Design in Zhou et al[60]

Type	End Point	Study Setting	Stopping Criterion
Example 1: Binary	ORR, defined using RECIST, version 1.1.	Evaluate the efficacy of pembrolizumab in treating patients with small bowel adenocarcinoma	ORR ≤ 0.2
Example 2: Ordinal	CR, PR, PD, SD	Evaluate the efficacy of nivolumab in in patients with Hodgkin's lymphoma[54]	CR + PR ≤ 0.3 and CR ≤ 0.15
Example 3: Coprimary	ORR and PFS6	Evaluate the efficacy of trebananib administered at 15 mg/kg IV per week in patients with persistent or recurrent carcinoma of the endometrium[55]	ORR ≤ 0.1 and PFS6 ≤ 0.2
Example 4: Efficacy and toxicity	ORR and toxicity rate	Evaluate the safety and efficacy of lenalidomide in combination with rituximab; rituximab in patients with recurrent indolent nonfollicular lymphoma[56]	ORR ≤ 0.45 or toxicity rate ≥ 0.3

Abbreviations: CR, complete remission; IV, intravenously; ORR, objective response rate; PD, progressive disease; PFS6, progression-free survival rate at 6 months; PR, partial remission; SD, stable disease.

TABLE 4. Optimal Stopping Boundaries for the Trial Examples in Table 3, Given a Type I Error Rate of .1

Trial	Stop Trial If	No. of Patients Treated						
		10	15	20	25	30	35	40
Example 1*	No. of OR ≤	1	2	4	5	7	9	10
Example 2†	No. of CR ≤	0	1	3	4	5	7	9
	and No. of CR/PR ≤	2	3	5	8	10	13	16
Example 3‡	No. of OR ≤	0	1	2	3	4	5	7
	and No, of PFS6 ≤	1	2	4	5	7	9	12
Example 4§	No. of OR ≤	2	5	7	10	13	16	19
	and No. of toxicity ≥	5	6	8	9	10	11	12

NOTE. The stopping boundaries are optimized on the basis of the null hypotheses (H_0) and alternative hypotheses (H_1).

Abbreviations: CR, complete remission; OR, objective response; PFS6, progression-free survival rate at 6 months; PR, partial remission.

*H_0 Pr(OR) = 0.2 versus H_1 Pr(OR) = 0.4.

†H_0 Pr(CR) = 0.15, Pr(CR or PR) = 0.30 versus H_1 Pr(CR) = 0.25, Pr(CR or PR) = 0.35.

‡H_0 Pr(OR) = 0.1, Pr(PFS6) = 0.20, Pr(OR and PFS6) =0.05 versus H_1 Pr(OR) = 0.3, Pr(PFS6) = 0.35, Pr(OR and PFS6) = 0.15.

§H_0 Pr(OR) = 0.45, Pr(toxicity) = 0.3, Pr(OR and toxicity) = 0.15 versus H_1 Pr(OR) = 0.60, Pr(toxicity) = 0.2, Pr(OR and toxicity) = 0.18.

decision making.[61] The stopping boundaries of the TOP design also can be enumerated and included in the trial protocol.

Simon's two-stage, BOP2, and TOP designs focus on one-arm trials. Brown et al[62] provide a comprehensive review on phase II trials in oncology, including randomized trials. The research on model-assisted designs for randomized phase II trials has been limited, and the designs warrant additional investigation.

SOFTWARE

The aforementioned model-assisted designs can be easily implemented using a Windows desktop program or online Web applications (freely available at trialdesign.org). Each module has an intuitive graphic user interface and rich documents to help users navigate through the process. A phase I or II trial can be designed easily by the following three steps:

1. Specify the design parameters (eg, sample size, cohort size, target DLT rate).
2. Use the software to produce decision table, design diagram, and operating characteristics of the design. The software also generates sample texts and protocol template to facilitate the protocol write-up.
3. Use the design decision table to conduct the trial and make adaptive decisions (eg, dose escalation/stay/de-escalation or go/no-go).

DISCUSSION

One major barrier for the adaptation of novel adaptive designs is that these designs often are complicated to implement. Model-assisted designs provide an attractive approach to remove this barrier and reconcile the conflict of simplicity versus performance by seizing the best of the two worlds. Model-assisted designs offer the superior performance compared with the more complicated, model-based adaptive designs but, once designed, can be implemented in as simple a way as the conventional designs. Implementation is facilitated even more by freely available user-friendly software. The approach establishes a new KISS principle: keep it simple and smart!

As in all trial designs, the design parameters for model-assisted designs must be carefully chosen to reflect the clinical setting and the study objective. For the BOIN design, the target DLT rate can vary with the type of phase I studies. The maximum sample size must be realistic and attainable. For the BOP2 design, the choice of using a binary end point or coprimary end point, and their null and target rates, depends on which disease type is studied and how effective the standard of care is. The number of interim analyses should account for both design efficiency and the logistic complexity. The choice of designs parameters should be validated and carefully calibrated through extensive computer simulation to ensure that the design has desirable operating characteristics under a variety of scenarios. The design parameters, study conduct boundaries, and operating characteristics must be spelled out and listed in the study protocol.

AFFILIATIONS

[1]University of Texas MD Anderson Cancer Center, Houston, TX
[2]Smith Breast Center, Baylor College of Medicine, Houston, TX

CORRESPONDING AUTHOR

Ying Yuan, PhD, University of Texas MD Anderson Cancer Center, 1400 Pressler St, Unit 1411, Houston, TX 77030; e-mail: yyuan@mdanderson.org.

SUPPORT

Supported in part by Grants No. 1P50CA217685 and 5P50CA098258 from the National Cancer Institute (Y.Y.) and in part by Grant No. CA016672 from the National Cancer Institute and Grant No. RP160668 from the Cancer Prevention and Research Institute of Texas (J.J.L.).

AUTHOR CONTRIBUTIONS

Conception and design: All authors
Collection and assembly of data: Ying Yuan, J. Jack Lee
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Ying Yuan
Honoraria: Ono Pharmaceutical
Consulting or Advisory Role: Boehringer Ingelheim Pharmaceuticals, Salzman Group, Midas Medical Technologies

J. Jack Lee
Consulting or Advisory Role: AbbVie

No other potential conflicts of interest were reported.

REFERENCES

1. Kessel M: The problems with today's pharmaceutical business: An outsider's view. Nat Biotechnol 29:27-33, 2011
2. US Food and Drug Administration: Speech to the Regulatory Affairs Professionals Society (RAPS) 2017 regulatory conference. https://www.fda.gov/news-events/speeches-fda-officials/speech-regulatory-affairs-professionals-society-raps-2017-regulatory-conference-09112017
3. DiMasi JA, Grabowski HG, Hansen RW: Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ 47:20-33, 2016
4. US Food and Drug Administration: Guidance document: Adaptive designs for clinical trials of drugs and biologics. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics
5. Yuan Y, Nguyen HQ, Thall PF: Bayesian Designs for Phase I-II Clinical Trials. Boca Raton, FL, Chapman and Hall/CRC, 2016
6. Yin G: Clinical Trial Design: Bayesian and Frequentist Adaptive Methods. New York, NY, John Wiley & Sons, 2012
7. Carlin BP, Berry SM, Lee JJ, et al: Bayesian Adaptive Methods for Clinical Trials. Boca Raton, FL, CRC Press, 2010
8. Chevret S: Bayesian adaptive clinical trials: A dream for statisticians only? Stat Med 31:1002-1013, 2012
9. Rogatko A, Schoeneck D, Jonas W, et al: Translation of innovative designs into phase I trials. J Clin Oncol 25:4982-4986, 2007
10. Yuan Y, Zhou Y, Lee JJ: Early Phase Clinical Trials, in Mell et al (eds): Principles of Clinical Cancer Research, New York, NY, Springer, 2019
11. Zhou H, Murray TA, Pan H, et al: Comparative review of novel model-assisted designs for phase I clinical trials. Stat Med 37:2208-2222, 2018
12. Zhou H, Yuan Y, Nie L: Accuracy, safety, and reliability of novel phase I trial designs. Clin Cancer Res 24:4357-4364, 2018
13. Storer BE: Design and analysis of phase I clinical trials. Biometrics 45:925-937, 1989
14. Simon R, Freidlin B, Rubinstein L, et al: Accelerated titration designs for phase I clinical trials in oncology. J Natl Cancer Inst 89:1138-1147, 1997
15. Skolnik JM, Barrett JS, Jayaraman B, et al: Shortening the timeline of pediatric phase I trials: The rolling six design. J Clin Oncol 26:190-195, 2008
16. Le Tourneau C, Lee JJ, Siu LL: Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst 101:708-720, 2009
17. O'Quigley J, Pepe M, Fisher L: Continual reassessment method: A practical design for phase 1 clinical trials in cancer. Biometrics 46:33-48, 1990
18. Babb J, Rogatko A, Zacks S: Cancer phase I clinical trials: Efficient dose escalation with overdose control. Stat Med 17:1103-1120, 1998
19. Neuenschwander B, Branson M, Gsponer T: Critical aspects of the Bayesian approach to phase I cancer trials. Stat Med 27:2420-2439, 2008
20. Yin G, Yuan Y: Bayesian model averaging continual reassessment method in phase I clinical trials. J Am Stat Assoc 104:954-968, 2009
21. Iasonos A, Wilton AS, Riedel ER, et al: A comprehensive comparison of the continual reassessment method to the standard 3 + 3 dose escalation scheme in phase I dose-finding studies. Clin Trials 5:465-477, 2008
22. Yan F, Mandrekar SJ, Yuan Y: Keyboard: A novel Bayesian toxicity probability interval design for phase I clinical trials. Clin Cancer Res 23:3994-4003, 2017
23. Ji Y, Liu P, Li Y, et al: A modified toxicity probability interval method for dose-finding trials. Clin Trials 7:653-663, 2010
24. Guo W, Wang S-J, Yang S, et al: A Bayesian interval dose-finding design addressing Ockham's razor: mTPI-2. Contemp Clin Trials 58:23-33, 2017
25. Yuan Y, Hess KR, Hilsenbeck SG, et al: Bayesian optimal interval design: A simple and well-performing design for phase I oncology trials. Clin Cancer Res 22:4291-4301, 2016
26. Liu S, Yuan Y: Bayesian optimal interval designs for phase I clinical trials. J R Stat Soc Ser C Appl Stat 64:507-523, 2015
27. Ruppert AS, Shoben AB: Overall success rate of a safe and efficacious drug: Results using six phase 1 designs, each followed by standard phase 2 and 3 designs. Contemp Clin Trials Commun 12:40-50, 2018
28. Lin R, Yin G: Bayesian optimal interval design for dose finding in drug-combination trials. Stat Methods Med Res 26:2155-2167, 2017
29. Yuan Y, Lin R, Li D, et al: Time-to-event Bayesian optimal interval design to accelerate phase I trials. Clin Cancer Res 24:4921-4930, 2018
30. Mu R, Yuan Y, Xu J, et al: gBOIN: A unified model-assisted phase I trial design accounting for toxicity grades, and binary or continuous end points. J R Stat Soc Ser C Appl Stat 68:289-308, 2019
31. Lin R, Yin G: STEIN: A simple toxicity and efficacy interval design for seamless phase I/II clinical trials. Stat Med 36:4106-4120, 2017
32. Takeda K, Taguri M, Morita S: BOIN-ET: Bayesian optimal interval design for dose finding based on both efficacy and toxicity outcomes. Pharm Stat 17:383-395, 2018
33. Bernhardt MB, De Guzman MM, Grimes A, et al: Rapid infusion of rituximab is well tolerated in children with hematologic, oncologic, and rheumatologic disorders. Pediatr Blood Cancer 65:e26759, 2018
34. Place AE, Goldsmith K, Bourquin J-P, et al: Accelerating drug development in pediatric cancer: A novel phase I study design of venetoclax in relapsed/refractory malignancies. Future Oncol 14:2115-2129, 2018
35. Phan TG, Ma H, Lim R, et al: Phase 1 trial of amnion cell therapy for ischemic stroke. Front Neurol 9:198, 2018
36. Postel-Vinay S, Gomez-Roca C, Molife LR, et al: Phase I trials of molecularly targeted agents: Should we pay more attention to late toxicities? J Clin Oncol 29:1728-1735, 2011
37. June CH, Warshauer JT, Bluestone JA: Is autoimmunity the Achilles' heel of cancer immunotherapy? Nat Med 23:540-547, 2017
38. Cheung YK, Chappell R: Sequential designs for phase I clinical trials with late-onset toxicities. Biometrics 56:1177-1182, 2000
39. Liu S, Yin G, Yuan Y: Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity. Ann Appl Stat 7:1837-2457, 2013
40. Normolle D, Lawrence T: Designing dose-escalation trials with late-onset toxicities using the time-to-event continual reassessment method. J Clin Oncol 24:4426-4433, 2006
41. Lin R, Yuan Y: Time-to-event model-assisted designs to accelerate phase I clinical trials. arXiv preprint arXiv:1807.08393, 2018
42. Zhang L, Yuan Y: Designing Early-Phase Drug Combination Trials, in O'Quigley et al (eds): Handbook of Methods for Designing, Monitoring, and Analyzing Dose-Finding Trials. Boca Raton, FL, Chapman and Hall/CRC, 2017, p 123
43. Ying Y, Heng Z, Yanhong Z: Phase I cancer clinical trial design: Single and combination agents, in Peace et al (eds): Biopharmaceutical Applied Statistics Symposium, Volume 1. New York, NY, Springer, 2018

44. Yin G, Yuan Y: Bayesian dose finding in oncology for drug combinations by copula regression. J R Stat Soc Ser C Appl Stat 58:211-224, 2009
45. Yin G, Yuan Y: A latent contingency table approach to dose finding for combinations of two agents. Biometrics 65:866-875, 2009
46. Wages NA, Conaway MR, O'Quigley J: Continual reassessment method for partial ordering. Biometrics 67:1555-1563, 2011
47. Thall PF, Millikan RE, Mueller P, et al: Dose-finding with two agents in phase I oncology trials. Biometrics 59:487-496, 2003
48. Riviere MK, Yuan Y, Dubois F, et al: A Bayesian dose-finding design for drug combination clinical trials based on the logistic model. Pharm Stat 13:247-257, 2014
49. Cai C, Yuan Y, Ji Y: A Bayesian dose finding design for oncology clinical trials of combinational biological agents. J R Stat Soc Ser C Appl Stat 63:159-173, 2014
50. Braun TM, Wang S: A hierarchical Bayesian design for phase I trials of novel combinations of cancer therapeutic agents. Biometrics 66:805-812, 2010
51. Zhang L, Yuan Y: A practical Bayesian design to identify the maximum tolerated dose contour for drug combination trials. Stat Med 35:4924-4936, 2016
52. Pan H, Lin R, Yuan Y: Statistical properties of the keyboard design with extension to drug-combination trials. arXiv:1712.06718, 2017
53. Simon R: Optimal two-stage designs for phase II clinical trials. Control Clin Trials 10:1-10, 1989
54. Sacchi S, Marcheselli R, Bari A, et al: Safety and efficacy of lenalidomide in combination with rituximab in recurrent indolent non-follicular lymphoma: Final results of a phase II study conducted by the Fondazione Italiana Linfomi. Haematologica 101:e196-e199, 2016
55. Moore KN, Sill MW, Tenney ME, et al: A phase II trial of trebananib (AMG 386; IND#111071), a selective angiopoietin 1/2 neutralizing peptibody, in patients with persistent/recurrent carcinoma of the endometrium: An NRG/Gynecologic Oncology Group trial. Gynecol Oncol 138:513-518, 2015
56. Cheson BD, Pfistner B, Juweid ME, et al: Revised response criteria for malignant lymphoma. J Clin Oncol 25:579-586, 2007
57. Hobbs BP, Chen N, Lee JJ: Controlled multi-arm platform design using predictive probability. Stat Methods Med Res 27:65-78, 2018
58. Yuan Y, Guo B, Munsell M, et al: MIDAS: A practical Bayesian design for platform trials with molecularly targeted agents. Stat Med 35:3892-3906, 2016
59. Woodcock J, LaVange LM: Master protocols to study multiple therapies, multiple diseases, or both. N Engl J Med 377:62-70, 2017
60. Zhou H, Lee JJ, Yuan Y: BOP2: Bayesian optimal design for phase II clinical trials with simple and complex endpoints. Stat Med 36:3302-3314, 2017
61. Lin R, Coleman RL, Yuan Y: TOP: Time-to-event Bayesian optimal phase II trial design for cancer immunotherapy. J Natl Cancer Inst pii:djz049 [epub ahead of print on March 29, 2019].
62. Brown SR, Gregory WM, Twelves CJ, et al: A Practical Guide to Designing Phase II Trials in Oncology. New York, NY, John Wiley & Sons, 2014.

APPENDIX

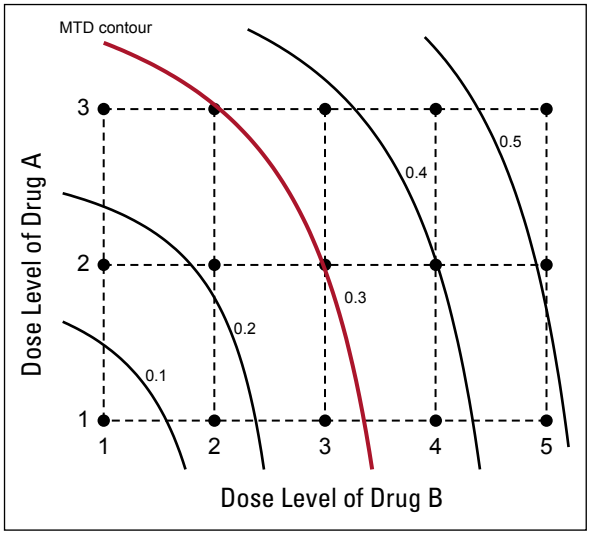


FIG A1. Maximum-tolerated dose (MTD) contour in drug combination trials. Curved lines indicated the toxicity contours with true toxicity rates of 0.1, 0.2, 0.3, 0.4, and 0.5, respectively. Combinations located along the rows and columns are ordered in toxicity, but, in other directions of the dose matrix (eg, along the diagonals from the upper left corner to the lower right corner), the toxicity order is unknown because of unknown drug-drug interactions.

