Statistics in Medicine

WILEY

Statistics in Medicine

RESEARCH ARTICLE OPEN ACCESS

UNITED: A Unified Transparent and Efficient Phase I/II Trial Design for Dose Optimization Accounting for Ordinal Graded, Continuous and Mixed Toxicity and Efficacy Endpoints

Mingyue Li[1] | Zhonglong Guo[2] | Yingjie Qiu[1,3]

[1]Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, Dallas, Texas, USA | [2]Co-Innovation Center for Sustainable Forestry in Southern China, College of Life Sciences, Nanjing Forestry University, Nanjing, China | [3]Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas, USA

Correspondence: Yingjie Qiu (yingjie.qiu@utsouthwestern.edu)

Received: 31 October 2024 | **Revised:** 5 March 2025 | **Accepted:** 6 April 2025

Funding: This work was supported by National Cancer Institute P30CA142543.

Keywords: adaptive design | delayed outcomes | efficacy grades | optimal biological dose | Phase I/II trial | toxicity grades

ABSTRACT

A primary objective of oncology dose-finding trials for novel therapies is to determine an optimal biological dose (OBD) that is both tolerable and therapeutically beneficial for patients in subsequent clinical trials. These new therapeutic agents are more likely to induce multiple low- or moderate-grade toxicities rather than dose-limiting toxicities. Additionally, efficacy is evaluated comprehensively, differentiating between complete remission and partial remission, as well as incorporating continuous efficacy endpoints. This important issue was highlighted in the American Statistical Association (ASA) Biopharmaceutical (BIOP) Section open forums and was a significant consideration of the FDA's "Project Optimus." We proposed the UNITED design, a unified, transparent, and efficient Phase I/II trial design to incorporate toxicity and efficacy grades and types, as well as continuous efficacy responses, into dose-finding and optimization. The UNITED design can handle binary, quasi-binary, continuous, and mixed toxicity and efficacy endpoints. We further extended the UNITED design, referred to as TITE-UNITED, to accommodate delayed toxicity and efficacy outcomes. Simulation studies showed that the UNITED and TITE-UNITED designs have desirable operating characteristics, performing comparably to or better than existing designs. A user-friendly software is available for practical implementation.

1 | Introduction

Traditionally, the primary objective of Phase I oncology trials has been to identify the maximum tolerated dose (MTD) based on dose-limiting toxicity (DLT). DLT was often defined as treatment-related grade 3 or higher non-hematologic toxicity or grade 4 or higher hematologic toxicity, according to the National Cancer Institute's Common Terminology Criteria for Adverse Events (CTCAE). This DLT-based dose-finding paradigm considered only the most severe toxicities in determining the tolerable dose, guiding decisions on dose escalation and de-escalation. Assuming toxicity data were summarized as a binary outcome, scored as either DLT or not DLT, numerous dose-finding methods have been proposed, such as the algorithm-based 3 + 3 design

[1], the model-based continual reassessment method [2], and the model-assisted Bayesian optimal interval (BOIN) design [3].

However, this conventional approach presents challenges for certain novel targeted therapies and immunotherapies. These treatments were often associated with multiple low- or moderate-grade toxicities that, while not severe enough to be classified as DLTs, could significantly impact patients' quality of life [4–6]. These lower-grade toxicities could lead to dose interruptions, discontinuations, and reductions, which might ultimately affect the efficacy of the therapies. Consequently, it is essential to incorporate various grades of toxicity into the dose-finding process to better assess and determine the MTD. This critical issue has been highlighted and discussed in the American Statistical Association (ASA) Biopharmaceutical (BIOP) Section Open Forums, organized by the ASA BIOP Statistical Methods in Oncology Scientific Working Group and the U.S. Food and Drug Administration (FDA) [7].

A number of approaches have been proposed to incorporate grades and types of toxicities into dose-finding trials. One general approach to incorporating toxicity grades into dose-finding is assigning a severity weight to each grade and type of toxicity event and combining the weights as a composite score. Bekele and Thall [8] proposed the total toxicity burden (TTB) as the arithmetic sum of different grades and types of toxicity, weighted by the severity weights elicited from clinicians. Similarly, Lee et al. [9] developed the toxicity burden score (TBS), which summarizes toxicity through a weighted sum, with severity weights estimated via regression using historical data. Further developments include the total toxicity profile (TTP), proposed by Ezzalfani et al. [10], computed as the Euclidean norm of severity weights. While these composite scores are not inherently mathematically continuous, appropriate transformations can approximate them as normally distributed endpoints. Yuan et al. [11], developed an equivalent toxicity score (ETS) approach that converts toxicity grades into numeric scores reflecting their relative severity in the unit of DLT, and developed the quasi-CRM to find MTD based on ETS. More recently, Mu et al. [12] introduced the generalized Bayesian optimal interval (gBOIN) design, which unifies various existing toxicity grade and type scoring systems within a single framework (e.g., TTB, TTP, TBS, ETS).

The conventional dose-finding designs based on toxicity pose challenges for novel anticancer agents, such as molecular targeted agents (MTAs), immunotherapy (IT), antibody-drug conjugates (ADCs), and chimeric antigen receptor (CAR) T-cell therapy. These agents often exhibit plateaued patterns in their dose-efficacy relationships, unlike traditional cytotoxic agents, for which both efficacy and toxicity monotonically increase with dose. For example, due to their mechanism of biological action, the dose-efficacy relationship may plateau at an intermediate dose when exposure achieves a saturation level in the body; thus, further increasing the dose level may not improve efficacy [13, 14]. In response to this issue, the FDA initiated Project Optimus to reform the dose selection optimization paradigm in oncology drug development [15]. The primary purpose of a dose-finding trial for novel anticancer agents is to identify an optimal biological dose (OBD), defined as the tolerable dose with adequate efficacy under unpredictable dose-toxicity and dose-efficacy relationships. Numerous Phase I/II designs have been suggested for identifying the OBD by incorporating both efficacy and toxicity responses in early-phase dose-finding trials [16–23]. Most of these designs assume that efficacy and toxicity data are summarized as binary outcomes, scored as a response or non-response for efficacy, and as a DLT or non-DLT for toxicity. However, it is crucial to incorporate the grades of toxicity, multiple levels of efficacy, and continuous efficacy responses, such as biomarker expression, white blood cell production, and pharmacodynamics (PD), in dose-finding and decision-making to determine the OBD for novel therapeutic agents [24–26] Takeda et al. [27] utilized the ETS approach proposed by Yuan et al. [11] and extended the ETS approach for efficacy grades as equivalent efficacy score (EES). With these efforts, a model-assisted approach that incorporates both efficacy and toxicity grades into the dose-finding process was proposed. Furthermore, the TITE-gBOIN-ET design, an extension of the gBOIN-ET design, was proposed as a model-assisted design aimed at accelerating early-phase trials by efficiently integrating efficacy and toxicity outcomes to shorten trial durations [28]. However, a comprehensive and effective framework that incorporates various types of toxicity and efficacy endpoints into dose-finding has not yet been proposed.

In this paper, we propose a *uni*fied *t*ransparent and *e*fficient Phase I/II trial design for *d*ose optimization (UNITED) to incorporate various clinical significant *t*oxicity and *e*fficacy endpoints, such as toxicity and efficacy grades, or continuous efficacy responses. Our design is a comprehensive extension of the mISO design for dose optimization based only on DLT and binary efficacy endpoints [22]. Specifically, we employ a quasi-binomial method to manage graded toxicity and efficacy outcomes by converting different toxicity and efficacy grades into an equivalent number of DLTs and responses, such as ETS and EES. Additionally, we discuss methods for addressing toxicity grades and types using composite scores such as TTB, TTP, TBS, and continuous efficacy endpoints. The UNITED design can also accommodate mixed types of toxicity and efficacy endpoints. To address the common issue of late-onset toxicity and efficacy in novel therapies, which may occur over multiple cycles, and the differences in the assessment window periods between efficacy and toxicity outcomes that could delay decision-making, we further extend the UNITED design to TITE-UNITED (time-to-event UNITED). This extension enables real-time decision-making while allowing sequential enrollment even when some patients' toxicity and efficacy assessments are pending.

The remainder of this article is structured as follows: In Section 2, we briefly review the mISO design. In Section 3, we introduce the UNITED design and present its statistical properties. Section 4 provides a comprehensive simulation study to evaluate the operating characteristics of UNITED and compare it with existing methodologies. Finally, in Section 5, we discuss our findings and conclude with final remarks. The software used to simulate and conduct the design is available at https://github.com/FrankQiu20.

2 | mSIO Design

We begin by briefly reviewing the mISO design, which aims to identify the OBD based on binary toxicity and efficacy endpoints. The mISO design uniquely allows for the simultaneous

monitoring of efficacy and toxicity outcomes within a single trial framework, effectively capturing the efficacy plateau of dose-efficacy relationships. Let $D = (d_1, \ldots, d_J)$ represent a set of J prespecified increasing dose levels under investigation. We denote p_{Tj} and p_{Ej} as the binary toxicity and efficacy rates, respectively, at each dose level $d_j, j = 1, \ldots, J$. The dose-response relationships based on the biological mechanism of the novel agents are captured to guide dose-finding design and algorithm. Specifically, for toxicity, a generally monotonically increasing dose-toxicity relationship is represented as

$$p_{T1} < \cdots < p_{TJ} \tag{1}$$

Then for efficacy, a plateaued dose-efficacy curve is specified as

$$p_{E1} < \cdots < p_{Ej^\dagger} = p_{Ej^\dagger+1} = \cdots p_{EJ} \tag{2}$$

Suppose at the current dose level j, y_{Tj}, and y_{Ej} out of n_j patients have experienced toxicity and efficacy respectively. A beta-binomial model is fitted for each dose level to construct an admissible set, which is used to exclude doses that are overly toxic or futile. Let $\mathcal{O}$ denote the observed data,

$$y_{Tj}|n_j, p_{Tj} \sim \text{Binomial}(n_j, p_{Tj}); \quad y_{Ej}|n_j, p_{Ej} \sim \text{Binomial}(n_j, p_{Ej}),$$
$$p_{Tj} \sim \text{Beta}(\alpha_T, \beta_T); \quad p_{E,j} \sim \text{Beta}(\alpha_E, \beta_E),$$
$$p_{Tj}|\mathcal{O} \sim \text{Beta}(\alpha_T + y_{Tj}, \beta_T + n_j - y_{Tj}); \quad p_{Ej}|\mathcal{O} \sim \text{Beta}(\alpha_E + y_{Ej}, \beta_E + n_j - y_{Ej}).$$

Let ϕ_T be the highest acceptable toxicity rate and ϕ_E be the lowest acceptable efficacy rate. By utilizing the non-decreasing dose-toxicity and dose-efficacy relationship specified at (1) and (2), the admissible set of toxicity $\mathcal{A}_T$ and efficacy $\mathcal{A}_E$ is constructed as follows:

$$d_{j_T} = \text{argmin}_D\big(pr(p_{Tj} > \phi_T|\mathcal{O}) > \theta\big); \quad \mathcal{A}_T = \{1 : d_{j_T-1}\}$$
$$d_{j_E} = \text{argmax}_D\big(pr(p_{Ej} < \phi_E|\mathcal{O}) > \xi\big); \quad \mathcal{A}_E = \{d_{j_E+1} : d_J\},$$

where θ and ξ represent the pre-determined cut-off probability for toxicity and efficacy respectively (e.g., 0.9 or 0.95). The OBD is defined as the lowest dose that yields the highest efficacy rate within the overall admissible set $\mathcal{A} = \mathcal{A}_T \cap \mathcal{A}_E$. Due to the plateau characteristic of the dose-efficacy relationship, it is common for multiple doses to meet this criterion. To minimize the risk of toxicity, the lowest dose among these is selected as the OBD.

To estimate p_{Ej}, the mISO design performs constrained nonparametric maximum likelihood estimation under the order restriction $p_{E1} < \cdots < p_{Ej^\dagger}$ based on the efficacy likelihood function:

$$L_E \propto \left\{\prod_{j=1}^{j^\dagger-1} p_{Ej}^{y_{Ej}}(1-p_{Ej})^{n_j-y_{Ej}}\right\} p_{Ej^\dagger}^{\sum_{j=j^\dagger}^{J} y_{Ej}}(1-p_{Ej^\dagger})^{\sum_{j=j^\dagger}^{J} n_j - y_{Ej}}.$$

The mISO design proposes the "collapse-then-split" algorithm, initially assuming a known location for $j^\dagger$, then relaxing this assumption, given the limited number of doses under investigation. This method uses the Pool-Adjacent-Violators Algorithm (PAVA) [29], known for its computational speed and transparency. The resulting isotonic estimates represent the maximum likelihood estimates (MLE) under the constraints of monotonically increasing order [29]. However, directly applying the PAVA proves inefficient due to the unique right-side plateau pattern. Algorithm 1 describes the details of the "collapse-then-split" approach, then all potential locations of $j^\dagger$ are enumerated. Model selection criteria based on the Akaike Information Criterion (AIC) are then utilized to derive the efficacy estimates and identify the location of the plateau point.

Dose escalation and de-escalation decisions within the mISO design are based on the admissible set, plateau location, and efficacy estimates. The dose-finding process is schematically represented in Figure 1. Given that the mISO design employs a non-parametric method that is capable only of estimating doses that have been tested, it incorporates a dose exploration rule to facilitate data collection in the trial. Specifically, if the highest tested dose, d_{j°, is deemed safe and d_{j° is not the highest dose level, the mISO design will escalate to the next higher dose.

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

1. Treat the first cohort of patients at the lowest dose d_1, or at the physician-specified dose.
2. Identify the highest tried dose d_{j°. If $\Pr(p_{Tj^\circ} > \phi_T|\mathcal{O}) \leq \theta$ and $j^\circ < J$, treat the next cohort of patients at dose $d_{j^\circ+1}$ and go to step 5; otherwise, go to step 3.
3. Construct the admissible set $\mathcal{A}$ using the proposed beta-binomial model. If $\mathcal{A}$ is empty, early terminate the trial and claim that no dose should be selected; otherwise, go to step 4.
4. Employ the proposed modified isotonic regression method to estimate the efficacy rates for all doses within $\mathcal{A}$. Identify the lowest dose level that yields the highest efficacy rate

ALGORITHM 1 | "Collapse-then-split" procedure.

Input: $\{y_{Ej}, n_j | \, j = 0, \ldots, J\}$

1: **Initialization**: Let $\boldsymbol{p^{(0)}} = (\hat{p}_{E1}, \ldots, \hat{p}_{Ej\triangle})$, and $\boldsymbol{w^{(0)}} = (w_1^{(0)}, \ldots, w_{j\triangle}^{(0)})$. $\hat{p}_{E1} = y_{E1}/n_1, \ldots, \hat{p}_{Ej\triangle} = \sum_{j=j^\dagger}^{J} y_{E,j} / \sum_{j=j^\dagger}^{J} n_j$, and $w_1^{(0)} = n_j, \ldots, w_{j\triangle}^{(0)} = \sum_{j=j^\dagger}^{J} n_j$.

2: **Requirement**: $\hat{p}_{E1} \leq \cdots \leq \hat{p}_{Ej\triangle}$

1. Satisfy - $\boldsymbol{p^{(0)}}$ are the final solutions
2. Not satisfy - apply PAVA to compute $\boldsymbol{p^{(1)}} = \{\hat{p}_{E1}^{(1)}, \ldots, \hat{p}_{E,j\triangle}^{(1)}\}$ using $\boldsymbol{p^{(0)}}$ with associated weights $\boldsymbol{w^{(0)}}$

3: **Finalization:** Re-define the isotonic estimates in plateau as $\hat{p}_{Ej^\dagger}^{(1)} = \cdots = \hat{p}_{EJ}^{(1)} = \hat{p}_{E,j\triangle}^{(1)}$

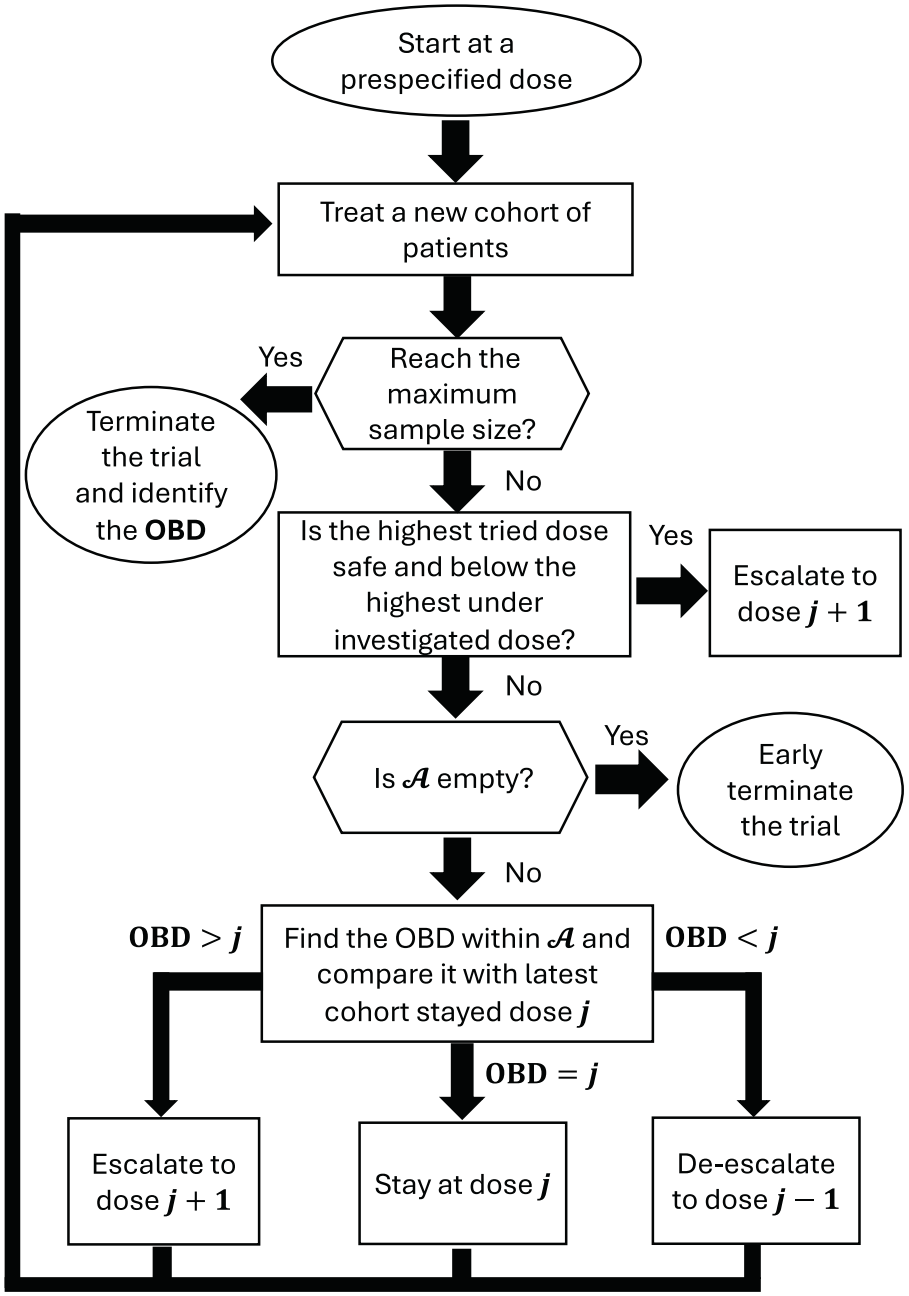

FIGURE 1 | Illustration of mISO dose-finding design.

estimate as the OBD. Let d_j represent the current dose level: if OBD $< d_j$, treat the next cohort of patients at dose d_{j-1}; if OBD $= d_j$, treat the next cohort of patients at dose d_j; if OBD $> d_j$, treat the next cohort of patients at dose d_{j+1}.

5. Repeat steps 2–4 until the maximum sample size has been reached.

At the end of the trial, the final admissible set is constructed, and efficacy rates for all doses within this set are estimated using the accumulated data. The lowest dose that yields the highest efficacy rate estimate within the admissible set is recommended as the final OBD.

3 | UNITED Design

The mISO design requires binary outcomes for both toxicity and efficacy. However, to optimize dose-finding and decision-making for identifying the OBD in the development of therapeutic agents, it's important to incorporate grades and types of toxicity, multiple levels of efficacy, and continuous efficacy outcomes. Therefore, the mISO design serves as the foundational basis for our extended comprehensive framework, the UNITED design, which is equipped to handle ordinal graded, continuous, and mixed toxicity and efficacy endpoints. Furthermore, with delayed toxicity and/or efficacy, some patients may have pending outcomes, rendering the direct application of the UNITED decision rule inefficient and infeasible. To address this challenge, we propose

an extension of the UNITED framework to the TITE-UNITED design, which accommodates delayed outcomes and refines decision-making processes.

3.1 | Quasi-Binary Toxicity and Efficacy Endpoints

An effective method for addressing ordinal graded toxicity and efficacy is the use of equivalent toxicity score (ETS) and equivalent efficacy score (EES) frameworks [11, 27]. These frameworks convert various grades of toxicity and efficacy into equivalent numbers of dose-limiting toxicities (DLTs) and responses, respectively. For example, we define that grade 0 and 1 toxicities are equivalent to 0 DLT, a grade 2 toxicity is equivalent to 0.5 DLT, a grade 3 toxicity is equivalent to 1 DLT, and a grade 4 toxicity is equivalent to 1.5 DLTs. Similarly, stable disease (SD) and progressive disease (PD) are considered equivalent to 0 response, partial response (PR) to 1 response, and complete response (CR) to 1.5 responses. We define the number of possible ordinal grades for toxicity and efficacy as $G_T + 1$ and $G_E + 1$, respectively. These grades range from 0 to G_T for toxicity, where G_T denotes the most severe adverse events, such as death, and from 0 to G_E for efficacy, where G_E corresponds to the most desirable efficacy response, such as CR. Let $\boldsymbol{w_M} = (w_{M0}, w_{M1}, \ldots, w_{MG_M})$ denote the equivalent number of DLTs or responses ascribed to grade $(0, 1, \ldots, G_M)$ for $M \in \{T, E\}$. Let $\boldsymbol{y_{Mj}} = (y_{Mj0}, \ldots, y_{MjG_M})$ denote the observed counts of grade g toxicities or efficacy outcomes among n_j patients treated at dose level d_j, where $g = 0, \ldots, G_M$. Here we make an assumption that for each patient, only one type of toxicity grade that a patient experienced is counted when defining y_{Tj}. For example, if a patient experience a grade 1 dermatitis and a grade 2 nausea, only grade 2 nausea will be counted. In instances where it is necessary to account for all grades of toxicity a patient experiences, the TTB approach is utilized. This method will be elaborated in the following section. Then, the total number of (equivalent) DLTs and responses (i.e., ETS and EES) at dose level d_j is given by

$$r_{Mj} = \boldsymbol{w_M} \boldsymbol{y_{Mj}^T}.$$

As an example, consider toxicities with $G_T = 4$ grades and $\boldsymbol{w_T} = (0, 0, 0.5, 1, 1.5)$ and efficacy with $G_E = 2$ grades and $\boldsymbol{w_E} = (0, 1, 1.5)$, suppose that at dose d_j three patients have been treated with three grade 2 toxicities were observed, then $r_{Tj} = 3 \times 0.5 = 1.5$. Suppose two PR and one CR were observed, $r_{Ej} = 2 \times 1 + 1 \times 1.5 = 3.5$ can be derived in the same way. For comparison, if the standard definition of DLT (e.g., grade 3 or higher toxicity) and objective response rate (ORR) (e.g., PR and CR) is used as the endpoint, we will completely ignore the graded 2 toxicity and partially ignore the better CR outcome.

It is important to note that the ETS and EES are not necessarily integers and are not bounded by n_j, thus the beta-binomial model and the "collapse-then-split" algorithm cannot be directly used. Following Yuan et al. [11] and Takeda et al. [27] and quasi-likelihood theory, we address those issue by standardizing r_{Mj} as r^*_{Mj} to $[0, n_j]$ using the quasi-binomial method:

$$r^*_{Mj} = \frac{r_{Mj}}{w_{MG_M}}; \quad r^*_{Mj} | n_j, p_{Mj} \sim \text{Quasi-Binomial}(n_j, p_{Mj}),$$

where p_{Mj} represents the quasi-binary toxicity and efficacy probabilities. As r^*_{Mj} can be modeled as a binomial distribution, the aforementioned beta-binomial model for admissible sets and constrained nonparametric maximum likelihood estimation under the order restrict $r^*_{E1} < \cdots < r^*_{Ej^\dagger}$ can still be applied as

$$p_{Tj} | \mathcal{O} \sim \text{Beta}(\alpha_T + r^*_{Tj}, \beta_T + n_j - r^*_{Tj}); \quad p_{Ej} | \mathcal{O} \sim \text{Beta}(\alpha_E + r^*_{Ej}, \beta_E + n_j - r^*_{Ej})$$

$$L_E \propto \{\prod_{j=1}^{j^\dagger - 1} p_{Ej}^{r^*_{Ej}} (1 - p_{Ej})^{n_j - r^*_{Ej}}\} p_{Ej^\dagger}^{\sum_{j=j^\dagger}^{J} r^*_{Ej}} (1 - p_{Ej^\dagger})^{\sum_{j=j^\dagger}^{J} n_j - r^*_{Ej}}.$$

Therefore, the "collapse-then-split" algorithm in Algorithm 1 and model selection criteria based on AIC can be adopted to handle quasi-binary toxicity and efficacy endpoints. Using the same dose-finding and optimization rules as mISO, the UNITED design can be easily applied. One advantage of using the ETS and EES is that they provide interpretations similar to the standard definitions of DLT and ORR. To draw the ϕ_T and ϕ_E, with the prespecificed target ETS and EES, ϕ_T = target ETS/$\text{ETS}_{\max}$ and ϕ_E = target EES/$\text{EES}_{\max}$, where $\text{ETS}_{\max}$ is the ETS for the most severe toxicity grade G_T, for example, grade 4, and $\text{EES}_{\max}$ is the EES for the most desirable efficacy response G_E. This approach ensures that ϕ_T and ϕ_E are scaled appropriately. Yuan et al. [11] consider the target ETS derived from the target toxicity profile at the target MTD: 49% grade 0 and grade 1 toxicity, 18% grade 2 toxicity, 23% grade 3 toxicity, and 10% grade 4 toxicity. Given the equivalent score $\boldsymbol{w_T}$, the target ETS is obtained by the weighted sum as $0.49 \times 0 + 0.18 \times 0.5 + 0.23 \times 1.0 + 0.1 \times 1.5 = 0.47$, and the highest acceptable $\phi_T = 0.47/1.5 = 0.313$. Following the rationale for specifying the target EES as detailed in gBOIN-ET design [27], we consider the lowest acceptable efficacy profile based on recent clinical trial results for primary peritoneal cancer [30]. This profile assumes 50% SD and PD, 30% PR, and 20% CR responses. As a result, the $\phi_E = (0.5 \times 0 + 0.3 \times 1.0 + 0.2 \times 1.5)/1.5 = 0.4$. It is crucial to elicit the target ETS and EES from both clinical and statistical considerations. A notable approach is proposed by Chen et al. [31], which involves a comprehensive toxicity scoring system measuring the composite severity of multiple toxicities. This system provides an effective solution for the common cases of multiple toxicities per patient.

3.2 | Continuous Toxicity and Efficacy Endpoints

An alternative method for quantifying toxicity incorporates composite scores such as TTB, TTP, and TBS. For clarity, we focus primarily on TTB, although the proposed methodology is equally applicable to other composite score endpoints like TTP and TBS. We adopt a similar example in Table 1 as shown by MT-keys design [7] for Phase I clinical trials. Consider an example where a patient experiences grade 4 Dermatitis, grade 3 Nausea or vomiting, and grade 4 Fatigue, the TTB for that patient is $6.0 + 1.5 + 1.0 = 8.5$. Although TTB is technically not a continuous variable due to the fixed number of types and grades of toxicities, it can often be well approximated as a continuous variable through appropriate transformation. Previous research has shown that ordinal variables with five or more categories can be treated as continuous variables with minimal impact on the analysis

TABLE 1 | Severity weights in the sarcoma trial.

	Type of toxicity	Grade	Severity weight
1	Myelosuppression without fever	3	1.0
		4	1.5
	Myelosuppression with fever	3	5.0
		4	6.0
2	Dermatitis	3	2.5
		4	6.0
3	Liver	2	2.0
		3	3.0
		4	6.0
4	Nausea/vomiting	3	1.5
		4	2.0
5	Fatigue	3	0.5
		4	1.0

[7, 32]. For efficacy, we consider endpoints that are continuous. Examples of such endpoints include biomarker expression levels, the production of white blood cells, and pharmacodynamic (PD) [24, 26, 33, 34]. Building on this foundation, we extend the UNITED design to accommodate continuous toxicity and efficacy endpoints.

Denote y_{Mji} as TTB or efficacy measurements for ith patient treated at dose level d_j, where $i = 1, \ldots, n_j$, and $M \in \{T, E\}$. We assume that y_{Mji} follows a normal distribution:

$$y_{Mji} \sim N(\mu_{Mj}, \sigma^2_{Mj}),$$

where μ_{Mj} and σ^2_{Mj} are mean and variance, respectively. Assuming a non-informative prior $p(\mu_{Mj}, \sigma^2_{Mj}) \propto \sigma^2_{Mj}$, given the observed data $\mathcal{O}$, the posterior distribution of μ_{Mj} follows a t-distribution with mean $\hat{\mu}_{Mj}$, scale s^2_{Mj} and $n_j - 1$ degrees of freedom,

$$\mu_{Mj}|\mathcal{O} \sim t_{n_j-1}(\hat{\mu}_{Mj}, \frac{s_{Mj}}{\sqrt{n_j}}),$$

where $\hat{\mu}_{Mj} = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{Mji}$, and $s^2_{Mj} = \frac{\sum_{i=1}^{n_j}(y_{Mji}-\hat{\mu}_{Mj})^2}{n_j-1}$. The detailed proof can be found in the Supporting Information. The posterior distributions are utilized to construct the admissible sets for toxicity and efficacy.

A significant challenge in extending the UNITED design to accommodate continuous endpoints is estimating efficacy through nonparametric constrained maximum likelihood estimation. To the best of our knowledge, there is no existing literature that addresses plateau-based nonparametric estimation without assuming an underlying curve. We address this issue and the likelihood function is given by

$$L_E \propto \left\{\prod_{j=1}^{j^\dagger-1}(\sigma^2_{Ej})^{-\frac{n_j}{2}} \exp\left(-\frac{\sum_{i=1}^{n_j}(y_{Eji}-\mu_{Ej})^2}{2\sigma^2_{Ej}}\right)\right\} \times \left\{\prod_{j=j^\dagger}^{J}(\sigma^2_{Ej})^{-\frac{n_j}{2}} \exp\left(-\frac{\sum_{i=1}^{n_j}(y_{Eji}-\mu_{Ej^\dagger})^2}{2\sigma^2_{Ej}}\right)\right\} \quad (3)$$

The likelihood function (3) depends on the parameters $\boldsymbol{\mu_E} = (\mu_{E1}, \ldots, \mu_{Ej^\dagger})$, and $\boldsymbol{\sigma^2_E} = (\sigma^2_{E1}, \ldots, \sigma^2_{Ej^\dagger}, \ldots, \sigma^2_{EJ})$. By assuming $j^\dagger$ is known, we perform constrained maximum likelihood estimation under the order constraints on the mean parameters $\mu_{E1} < \cdots < \mu_{Ej^\dagger}$, without imposing any assumptions or constraints on the variances σ^2_{Ej}. We adapt the two-step iterative algorithm based on the Pool-Adjacent-Violators Algorithm (PAVA), as proposed by Shi and Jiang [35] and propose the "Collapse-then-split" two-step iterative algorithm in Algorithm 2. This approach is to maximize the logarithm of the likelihood function (3), with mean vector $\boldsymbol{\mu} \in D$, and $\boldsymbol{\sigma^2_E} \in \mathbb{R}^J_+$, where D is the set of isotonic vector in $\mathbb{R}^{j^\dagger}$. The development of the algorithm for estimation rely on Theorem 1 and 2. The detailed proof of Theorem 1 and 2 is attached in the Supporting Information.

Theorem 1. *MLE of $(\boldsymbol{\mu_E}, \boldsymbol{\sigma^2_E})$ under the order restrictions exists since a maximize of over L_E in (3) over D exists.*

Theorem 2. *The point sequence $\{\boldsymbol{\mu}^{(k)}\}$ given the Algorithm 2 converges to a maximizer of L_E in (3) as $k \to \infty$.*

ALGORITHM 2 | "Collapse-then-split" two steps iterative procedure.

Input: $\{y_{Eji}, n_j | \ j = 0, \ldots, J, \ i = 1, \ldots, n_j\}$

1: **Initialization**: Let $\boldsymbol{\mu^{(0)}} = (\hat{\mu}_{E1}, \ldots, \hat{\mu}_{Ej^\dagger}, \ldots, \hat{\mu}_{EJ})$, $\boldsymbol{\sigma^{2(0)}} = (\hat{\sigma}^2_{E1}, \ldots, \hat{\sigma}^2_{EJ})$, and $\boldsymbol{w^{(0)}} = (w^{(0)}_1, \ldots, w^{(0)}_J)$. $\hat{\mu}_{E1} = \frac{\sum_{i=1}^{n_1} y_{E1i}}{n_1}, \ldots, \hat{\mu}_{E,j\triangle} = \frac{\sum_{j=j^\dagger}^{J}\sum_{i=1}^{n_j} y_{Eji}}{\sum_{j=j^\dagger}^{J} n_j}$, $\hat{\mu}_{Ej^\dagger} = \cdots = \hat{\mu}_{EJ} = \hat{\mu}_{E,j\triangle}$, $\hat{\sigma}^2_{Ej} = \frac{\sum_{i=1}^{n_j}(y_{Eji}-\hat{\mu}_{Ej})^2}{n_j}$, and $w^{(0)}_j = \frac{n_j}{\sigma^{2(0)}}$.

2: **Requirement**: $\hat{\mu}_{E1} \le \cdots \le \hat{\mu}_{Ej^\dagger}$

1. Satisfy - $\boldsymbol{\mu^{(0)}}$ and $\boldsymbol{\sigma^{2(0)}}$ are the final solutions
2. Not satisfy - apply PAVA to compute $\boldsymbol{\mu^{(k)}}$ with associated weights $\boldsymbol{w^{(k-1)}}$
   (a) Compute $\boldsymbol{\sigma^{2(k)}} = (\hat{\sigma}^{2(k)}_{E1}, \ldots, \hat{\sigma}^{2(k)}_{EJ})$, and $\boldsymbol{w^{(k)}} = (w^{(k)}_1, \ldots, w^{(k)}_J)$, where $\hat{\sigma}^{2(k)}_{Ej} = \frac{\sum_{i=1}^{n_j}(y_{Eji}-\hat{\mu}^{(k)}_{Ej})^2}{n_j}$, and $w^{(k)}_j = \frac{n_j}{\sigma^{2(k)}}$
   (b) Check the order requirement using updated weight $\boldsymbol{w^{(k)}}$, repeat until

$$\overset{max}{1 \le j \le J} \left|\hat{\mu}^{(k-1)}_{Ej} - \hat{\mu}^{(k)}_{Ej}\right| \le 10^{-3}$$

By integrating admissible sets with efficacy estimation for continuous endpoints, we extend the mISO framework to UNITED, maintaining the same dose-finding design and algorithm. The UNITED design has been adapted to accommodate a range of endpoint types, including binary, quasi-binary, and continuous. As toxicity and efficacy are modeled independently within this framework, UNITED can also transparently handle mixed endpoints. Examples of UNITED's flexibility include scenarios with continuous toxicity and quasi-binary efficacy, as well as quasi-binary toxicity combined with continuous efficacy. These specific scenarios are detailed further in Section 4.4.3.

3.3 | Delayed Outcomes

One common challenge in the implementation of targeted therapies and immunotherapies is the late onset of toxicity and efficacy, which can manifest over multiple treatment cycles. Unlike traditional cytotoxic agents, which typically require assessment within the first 1–2 treatment cycles, delayed outcomes necessitate a longer evaluation window. This issue poses a significant challenge for the UNITED design, which relies on completely observed data. To address this, we propose the TITE-UNITED design using the approximated likelihood approach developed by Lin et al. [36]. While our focus is on quasi-binary endpoints like ETS and EES, it is important to note that this quasi-binomial approximated likelihood approach is also applicable to composite scores such as TTB. We presented the rationale and details in the Supporting Information.

Let $\tilde{r}^*_{Mji}, i = 1, \ldots, n_j$ be the observed quasi-binary toxicity or efficacy endpoints at dose level d_j, which indicates the patient i has experienced toxicity or efficacy ($0 < \tilde{r}^*_{Mji} \leq 1$), or not yet experienced ($\tilde{r}^*_{Mji} = 0$) at the time of dose assignment for the next incoming cohort. We should notice that $0 < \tilde{r}^*_{Mji} \leq 1$ implies $0 < r^*_{Mji} \leq 1$, but $\tilde{r}^*_{Mji} = 0$ implies $0 \leq r^*_{Mji} \leq 1$. It is because that a pending patient who has not experienced the event by the interim time may experience the event later during follow-up. Let δ_{Mji} denote the status of toxicity and efficacy outcome by the interim decision time, that is, $\delta_{Mji} = 1$ suggests the outcome has been ascertained while $\delta_{Mji} = 0$ is still pending. Let τ_M be the assessment window for toxicity/efficacy. We denote $\upsilon_{Mji} (\upsilon_{Mji} \leq \tau_M)$ and t_{Mji} be the actual follow up time and time-to-event of toxicity/efficacy at dose level d_j for patient i respectively. Therefore, by assuming the time-to-event is uniformly distributed over the assessment window recommend by previous research [36, 37], for a patient i with delayed outcome at dose level d_j, the likelihood can be approximated by:

$$
\begin{aligned}
&\Pr(\tilde{r}^*_{Mji} = 0 | \delta_{Mji} = 0) \\
&= \Pr(\tilde{r}^*_{Mji} = 0, r^*_{Mji} = 0) + \Pr(\tilde{r}^*_{Mji} = 0, r^*_{Mji} > 0) \\
&= \Pr(\tilde{r}^*_{Mji} = 0) + \Pr(r^*_{Mji} > 0) \Pr(t_{Mji} > \upsilon_{Mji} | r^*_{Mji} > 0) \\
&\approx 1 - p_{Mj} + p_{Mj} \frac{\tau_M - \upsilon_{Mji}}{\tau_M} = 1 - \frac{\upsilon_{Mji}}{\tau_M} p_{Mj} \approx (1 - p_{Mj})^{\frac{\upsilon_{Mji}}{\tau_M}}
\end{aligned}
$$

Thus, in the case of delayed outcome, the likelihood of toxicity/efficacy at dose level d_j is given by:

$$
\begin{aligned}
L(p_{Mj}) &\propto \prod_{i=1}^{n_j} p_{Mj}^{\delta_{Mji} r^*_{Mji}} (1 - p_{Mj})^{\delta_{Mji}(1 - r^*_{Mji})} (1 - p_{Mj})^{(1 - \delta_{Mji}) \frac{\upsilon_{Mji}}{\tau_M}} \\
&= p_{Mj}^{\tilde{r}^*_{Mj}} (1 - p_{Mj})^{\tilde{n}_j - \tilde{r}^*_{Mj}},
\end{aligned}
$$

where $\tilde{r}^*_{Mj} = \sum_{i=1}^{n_j} \delta_{Mji} r^*_{Mji}$ is the total observed quasi-binary ETS/EES outcomes by the interim decision time, and $\tilde{n}_j = \sum_{i=1}^{n_j} \delta_{Mji} + \sum_{i=1}^{n_j} (1 - \delta_{Mji}) \frac{\upsilon_{Mji}}{\tau_M}$ is the "effective" sample size of ETS/EES to accommodate the incomplete follow-up. It represents that at dose level d_j, the number of patients who have completed the assessment for toxicity/efficacy + (total follow-up time for pending patients for toxicity/efficacy) / (length of assessment window for toxicity/efficacy). This approach simplifies the complex irregular likelihood arising from a mixture of observed and pending outcomes. As a result, the aforementioned beta-binomial models and estimation methods of efficacy can be directly applied, potentially shortening the trial duration. For example, by utilizing effective efficacy data $\tilde{n}_j$ and $\tilde{r}^*_{Ej}$, the posterior beta distribution of p_{Ej} becomes $p_{Ej} | \mathcal{O} \sim \text{Beta}(\alpha_E + \tilde{r}^*_{Ej}, \beta_E + \tilde{n}_j - \tilde{r}^*_{Ej})$. Similarly, r^*_{Ej} is revised to $\tilde{r}^*_{Ej}$, and n_j to $\tilde{n}_j$, reflecting these adjustments in the likelihood function L_E described in Section 3.1. However, if the patient accrual is much faster than the outcome evaluation, we may not have sufficient information to make robust decisions although we use the pending data. To address this, we propose an enrollment suspension rule to ensure adequate information collection. For TITE-UNITED design, we require that dose assignment is not allowed until at least 50% of patients at the current dose level have completed both the toxicity and efficacy assessment.

4 | Numerical Studies

4.1 | Quasi-Binary Toxicity and Efficacy Endpoints

We conducted a simulation study to evaluate the operating characteristics (OCs) of the UNITED design, utilizing ETS and EES as quasi-binary endpoints. We considered a Phase I/II trial with 6 doses and a maximum sample size of 60 in cohorts of size 3. We investigated the commonly considered highest acceptable $\phi_T = 0.313$ (based on the highest acceptable ETS = 0.47) and the lowest acceptable efficacy $\phi_E = 0.4$ (based on the lowest acceptable EES = 0.6). We set $\alpha_T = \beta_T = \alpha_E = \beta_E = 1$ in the beta-binomial model to specify non-informative priors. We also specified the cut-off probability values $\theta = \xi = 0.9$, which were determined based on preliminary simulation studies that calibrated Bayesian design parameters. Nine scenarios were considered, each featuring different locations of the OBD and various shapes of dose-response curves. We compare the UNITED design with gBOIN-ET design [27], the gBOIN design [12], and the quasi-CRM design [11]. Since the gBOIN design and quasi-CRM design is primarily designed for determining the MTD, we adopted the two stage design (Phase 1 followed by cohort expansion). The same basic settings for the simulations were used in all four designs for a fair comparison with respect to maximum number of patients per trial (60), cohort size (3), starting

dose (lowest dose level), the highest acceptable standardized ETS (i.e., $ETS^S = 0.313$) and lowest acceptable standardized EES (i.e., $EES^S = 0.4$), and simulation times (10,000). The detailed simulation configuration was illustrated in the Supporting Information.

Table 2 summarized the operating characteristics of the four designs. The results were presented as the percentage of simulation times each design selected a particular dose level as the OBD, denoted as Selection (%), and the average percentage of patients treated at each dose level, referred to as Patient (%). Outcomes that reflected correct OBD selection were highlighted in boldface. The selection percentage under the dose level '0' indicated the percentage of early trial termination without identifying an OBD.

Under Scenarios 1–5, the MTD defined by the ETS and the OBD defined by both the ETS and EES were not matched. Specifically, in scenario 1, where the MTD was located at dose level 4 and the OBD was defined at dose level 1, the percentage of correct OBD selection of UNITED was 71.6%, which was 12.5%, 37.4%, and 36.2% higher than that of gBOIN-ET, gBOIN, and quasi-CRM. The percentage of patients treated at OBD of UNITED was 47.2%, which was slightly less than that of gBOIN-ET but significantly higher than that of gBOIN and quasi-CRM. Similar results were observed in the subsequent Scenarios 2–5.

Under Scenarios 6–8, the MTD defined by the ETS and the OBD defined by both the ETS and EES were matched. The UNITED design successfully identified the correct OBDs and allocated patients at percentages comparable to those of other designs. In Scenario 9, where no dose satisfied the toxicity and efficacy requirements, UNITED terminated the trial early in 96.1% of simulations. This rate was comparable to gBOIN-ET and significantly higher than that of gBOIN and quasi-CRM. Additionally, UNITED used the fewest average number of patients in all scenarios, contributing to significant savings in patient resources.

4.2 | Continuous Toxicity and Efficacy Endpoints

We utilized the TTB as an example of a continuous toxicity endpoint, in addition to continuous efficacy endpoints, to compare the UNITED design with gBOIN. Adopting the setting of the sarcoma trial [8], we considered six types of toxicities with different severity weights assigned to different toxicity grades as shown in Table 1. We considered 9 scenarios with varying locations of the OBD. The toxicity upper bound TTB was $\phi_T = 3.04$, and the efficacy lower bound $\phi_E = 0.3$. The maximum sample size was 60 patients and the cohort size was 3. The details of the nine scenarios (e.g., the probability of each type and grade of toxicities), and simulation configuration were provided in Supporting Information.

Table 3 summarized the results. The UNITED design demonstrated superior performance in terms of the correct selection of OBD and patient allocation when the OBD and the MTD were not matched, as illustrated in Scenarios 1–5. For example, in Scenario 2, the MTD was dose level 4, and the OBD was dose level 2 since it was safe and the efficacy reached the plateau from this dose. The percentage of correct selection of OBD and the percentage of patients treated at OBD of UNITED is 73.9% and 44.5% respectively, which was 43.1% and 39.1% higher than that of the gBOIN design. When the MTD and OBD matched, as in scenarios 6-8, gBOIN slightly outperformed UNITED. In scenario 9, where no dose met both acceptable toxicity and efficacy profiles, the percentages of early trial termination for the two designs were similar; however, UNITED was able to save almost half the number of patients compared to gBOIN.

4.3 | Delayed Outcomes

In our investigation of delayed outcomes, patient accrual was simulated from a uniform distribution, with an accrual rate of three patients per month. The toxicity and efficacy assessment windows were set at 1.5 months and 3 months, respectively, to reflect realistic trial settings. The time-to-efficacy and time-to-toxicity outcomes were simulated using a Weibull distribution, ensuring that 50% of the efficacy and toxicity events occurred in the latter half of their respective assessment windows. We compared the proposed TITE-UNITED design with the conventional UNITED design. The conventional approach suspends the trial and defers interim analysis if any pending outcomes exist among patients. While this method ensures comprehensive data collection, it may not be feasible in practice due to the potential for excessively long trial durations. Nonetheless, it serves as a useful reference for comparison.

Table 4 summarized the results. We reported the same OCs using the same ETS^S and EES^S across all nine scenarios, as previously detailed in Table 2. Besides, we also reported the trial duration under each design. The TITE-UNITED yielded slightly lower true correct OBD selection and patients' allocation percentages at OBD than the conventional UNITED design. However, the TITE-UNITED design significantly reduced the trial duration. For example, in scenario 3, the TITE-UNITED design achieved a correct OBD selection percentage of 76.9% and a patient allocation at OBD percentage of 40.2%, which were only 1.3% and 2.7% lower, respectively, than those under the conventional UNITED design. However, it reduced the trial duration from 70.1 months to 38.1 months. A similar pattern was observed in other scenarios. Therefore, when considering all design metrics, including OBD selection, patient allocation, and trial duration, the proposed TITE-UNITED design proves to be preferable to the conventional UNITED approach, particularly in scenarios involving delayed outcomes. In Table S4, we compare the performance of the TITE-UNITED design with the TITE-gBOIN-ET design [28]. Overall, the TITE-UNITED design shows improved performance in scenarios where the MTD and the OBD do not align, while exhibiting slightly lower performance in scenarios where the MTD and OBD are matched. In the TITE-UNITED design, we assumed that toxicity and efficacy grades were exclusive of each other. However, patients may experience multiple toxicity and efficacy events during the assessment period. For example, a patient may exhibit an early-on low-grade or mild response, followed by higher-grade toxicity or efficacy later. To evaluate the robustness of the TITE-UNITED design under these conditions, we conducted a sensitivity analysis that allowed for multiple toxicity and efficacy events per patient. Following a method similar to that used by Takeda et al., we sequentially generated the time-to-toxicity from the Weibull distribution [28]. If the time to the first toxicity outcome of Grade 2, 3, or 4 occurred

TABLE 2 | The operating characteristics (OCs) of the UNITED, gBOIN-ET, gBOIN and quasi-CRM designs including percentages of OBD selection and patients' allocation and number of patients in the trial, based on 10,000 simulated trials.

		Dose level							# of
Design		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**patients**
Scenario 1									
	True ETS^S		0.08	0.12	0.22	0.32	0.5	0.57	
	True EES^S		0.45	0.45	0.45	0.45	0.48	0.48	
UNITED	Selection (%)	5.5	**71.6**	14.8	4.9	3.1	0.1	0.0	
	Patient (%)		47.2	17.4	12.7	11.5	8.1	3.2	58.1
gBOIN-ET	Selection (%)	0.6	**59.1**	27.1	10.3	2.8	0.1	0.0	
	Patient (%)		63.5	25.8	8.2	2.2	0.3	0.0	59.7
gBOIN	Selection (%)	0.6	**34.2**	28.9	22.7	13.3	0.4	0.0	
	Patient (%)		5.8	10.2	31.6	41.1	10.5	0.9	60.0
quasi-CRM	Selection (%)	0.4	**35.4**	28.0	22.5	13.2	0.5	0.0	
	Patient (%)		6.6	10.0	31.0	45.4	6.9	0.0	60.0
Scenario 2									
	True ETS^S		0.05	0.14	0.21	0.33	0.42	0.53	
	True EES^S		0.30	0.50	0.50	0.51	0.51	0.51	
UNITED	Selection (%)	3.8	22.1	**65.0**	6.9	1.7	0.5	0.0	
	Patient (%)		27.3	37.2	13.3	10.4	8.3	3.6	58.7
gBOIN-ET	Selection (%)	0.0	11.3	**64.6**	21.0	2.8	0.3	0.0	
	Patient (%)		23.8	59.8	14.2	2.0	0.3	0.0	60.0
gBOIN	Selection (%)	0.1	11.8	**39.9**	30.9	15.4	1.8	0.0	
	Patient (%)		5.4	10.5	33.2	35.6	13.3	2.0	60.0
quasi-CRM	Selection (%)	0.1	12.6	**39.4**	30.1	16.0	1.8	0.0	
	Patient (%)		5.2	11.1	32.0	40.7	10.7	0.2	60.0
Scenario 3									
	True ETS^S		0.05	0.08	0.10	0.20	0.28	0.37	
	True EES^S		0.13	0.30	0.58	0.58	0.58	0.60	
UNITED	Selection (%)	0.5	0.9	14.6	**77.5**	5.4	1.0	0.1	
	Patient (%)		8.2	22.3	42.7	12.5	9.7	4.6	59.8
gBOIN-ET	Selection (%)	0.0	0.0	10.7	**76.5**	11.9	0.8	0.1	
	Patient (%)		6.0	22.1	60.7	10.7	0.5	0.0	60.0
gBOIN	Selection (%)	0.0	0.4	5.4	**39.0**	31.4	19.9	3.9	
	Patient (%)		5.2	5.6	8.2	24.9	32.9	23.3	60.0
quasi-CRM	Selection (%)	0.0	0.6	6.2	**41.8**	33.6	16.7	1.1	
	Patient (%)		5.9	5.7	7.4	31.3	43.2	6.6	60.0
Scenario 4									
	True ETS^S		0.05	0.08	0.10	0.12	0.20	0.37	
	True EES^S		0.07	0.13	0.20	0.45	0.45	0.46	
UNITED	Selection (%)	3.6	0.1	0.2	3.9	**78.0**	11.9	2.2	
	Patient (%)		5.8	8.9	18.1	44.7	15.8	6.7	58.9
gBOIN-ET	Selection (%)	0.0	0.1	0.2	0.8	**64.2**	30.3	4.3	
	Patient (%)		5.7	7.0	9.7	50.2	24.6	2.8	60.0
gBOIN	Selection (%)	0.7	0.5	2.1	4.8	**42.2**	36.5	13.1	
	Patient (%)		5.2	5.6	6.4	10.8	32.8	39.3	60.0

(Continues)

TABLE 2 | (Continued)

Design		Dose level 0	1	2	3	4	5	6	# of patients
quasi-CRM	Selection (%)	0.1	0.6	1.3	6.8	**48.2**	38.5	4.5	
	Patient (%)		5.9	5.7	6.2	13.7	54.1	14.5	60.0
Scenario 5									
	True ETS^S		0.05	0.08	0.10	0.12	0.13	0.25	
	True EES^S		0.03	0.07	0.13	0.22	0.46	0.46	
UNITED	Selection (%)	4.3	0.1	0.0	0.2	6.1	**81.4**	7.8	
	Patient (%)		5.4	6.2	10.3	21.9	47.2	9.0	58.7
gBOIN-ET	Selection (%)	0.0	0.1	0.1	0.5	2.7	**70.9**	25.7	
	Patient (%)		5.7	6.5	8.0	11.9	52.9	15.0	60.0
gBOIN	Selection (%)	0.7	0.3	1.0	2.7	9.3	**48.9**	37.1	
	Patient (%)		5.2	5.6	6.4	8.9	15.6	58.4	60.0
quasi-CRM	Selection (%)	0.3	0.4	2.0	5.1	10.5	**63.5**	18.2	
	Patient (%)		5.9	5.7	6.2	12.2	41.4	28.7	60.0
Scenario 6									
	True ETS^S		0.05	0.08	0.10	0.12	0.13	0.18	
	True EES^S		0.03	0.07	0.13	0.18	0.23	0.48	
UNITED	Selection (%)	9.4	1.5	0.0	0.2	1.8	13.2	**73.8**	
	Patient (%)		5.7	6.5	9.3	15.7	24.8	38.1	57.6
gBOIN-ET	Selection (%)	7.1	0.8	0.7	1.8	3.1	9.1	**77.4**	
	Patient (%)		6.1	6.8	8.3	10.9	17.3	50.5	58.8
gBOIN	Selection (%)	2.1	0.5	1.5	3.5	8.1	13.4	**70.9**	
	Patient (%)		5.2	5.6	6.4	8.9	12.2	61.7	60.0
quasi-CRM	Selection (%)	0.4	2.2	6.4	11.4	18.9	24.7	**36.0**	
	Patient (%)		5.9	5.7	6.2	12.2	40.4	29.6	60.0
Scenario 7									
	True ETS^S		0.10	0.27	0.39	0.50	0.59	0.77	
	True EES^S		0.27	0.53	0.55	0.63	0.73	0.73	
UNITED	Selection (%)	14.5	9.9	**70.3**	4.9	0.4	0.0	0.0	
	Patient (%)		26.3	50.0	12.9	6.6	3.2	1.0	55.6
gBOIN-ET	Selection (%)	12.1	24.7	**54.9**	6.1	1.4	0.7	0.1	
	Patient (%)		31.8	53.1	9.9	3.0	1.4	0.7	56.3
gBOIN	Selection (%)	2.1	10.6	**72.1**	14.3	1.0	0.0	0.0	
	Patient (%)		12.2	53.2	27.8	6.0	0.7	0.0	60.0
quasi-CRM	Selection (%)	0.3	13.2	**72.6**	13.6	0.3	0.0	0.0	
	Patient (%)		9.9	57.6	26.7	5.5	0.3	0.0	60.0
Scenario 8									
	True ETS^S		0.05	0.13	0.28	0.44	0.52	0.60	
	True EES^S		0.07	0.25	0.52	0.52	0.62	0.63	
UNITED	Selection (%)	14.7	0.3	9.6	**73.2**	2.2	0.1	0.0	
	Patient (%)		7.9	25.0	46.4	11.2	6.8	2.6	55.6
gBOIN-ET	Selection (%)	18.8	2.5	22.3	**51.7**	3.1	1.1	0.4	
	Patient (%)		8.3	30.6	49.7	7.4	2.7	1.3	55.5

(Continues)

TABLE 2 | (Continued)

Design		Dose level							# of patients
		0	**1**	**2**	**3**	**4**	**5**	**6**	
gBOIN	Selection (%)	3.3	0.3	10.1	**77.6**	8.7	0.1	0.0	
	Patient (%)		5.7	12.7	52.5	25.8	3.1	0.2	60.0
quasi-CRM	Selection (%)	0.8	0.9	14.3	**73.9**	10.1	0.0	0.0	
	Patient (%)		6.0	12.2	52.0	28.0	1.8	0.0	59.9
Scenario 9									
	True ETS^S		0.22	0.30	0.40	0.50	0.53	0.60	
	True EES^S		0.10	0.13	0.22	0.30	0.42	0.53	
UNITED	Selection (%)	**96.1**	0.0	0.1	1.5	2.0	0.3	0.0	
	Patient (%)		22.9	25.6	25.7	15.6	7.8	2.5	26.1
gBOIN-ET	Selection (%)	**93.3**	2.4	1.7	1.9	0.3	0.3	0.1	
	Patient (%)		34.0	26.0	20.5	11.1	5.9	2.4	28.8
gBOIN	Selection (%)	**79.1**	6.5	7.9	5.6	0.9	0.1	0.0	
	Patient (%)		31.4	44.0	20.0	4.0	0.6	0.0	59.9
quasi-CRM	Selection (%)	**76.6**	9.0	10.0	4.3	0.1	0.0	0.0	
	Patient (%)		39.5	41.9	15.8	2.7	0.1	0.0	59.2

TABLE 3 | The operating characteristics (OCs) of the UNITED, and gBOIN designs including percentages of OBD selection and patients' allocation and number of patients in the trial, based on 10,000 simulated trials.

Design		Dose level							# of patients
		0	**1**	**2**	**3**	**4**	**5**	**6**	
Scenario 1									
	True TTB		0.22	0.91	1.48	3.04	4.81	6.1	
	True efficacy		0.5	0.5	0.5	0.5	0.52	0.6	
UNITED	Selection (%)	0.0	**68.3**	19.2	8.6	3.9	0.0	0.0	
	Patient (%)		41.7	19.9	15.1	12.3	7.9	3.1	60.0
gBOIN	Selection (%)	0.0	**30.0**	30.1	26.3	13.6	0.1	0.0	
	Patient (%)		5.1	5.7	11.4	73.0	4.7	0.1	60.0
Scenario 2									
	True TTB		0.32	0.74	1.25	1.57	3.04	4.74	
	True efficacy		0.22	0.45	0.45	0.45	0.45	0.45	
UNITED	Selection (%)	0.0	0.0	**73.9**	18.1	6.1	1.9	0.1	
	Patient (%)		5.0	44.5	19.7	13.9	11.5	5.3	60.3
gBOIN	Selection (%)	0.0	0.0	**30.8**	28.9	25.6	14.5	0.2	
	Patient (%)		5.1	5.4	6.4	15.5	57.3	10.2	60.0
Scenario 3									
	True TTB		0.32	0.74	1.25	1.57	3.04	4.74	
	True efficacy		0.22	0.37	0.53	0.53	0.53	0.53	
UNITED	Selection (%)	0.0	0.0	0.1	**78.2**	16.5	5.1	0.1	
	Patient (%)		5.0	5.4	50.1	20.0	13.8	5.6	60.0
gBOIN	Selection (%)	0.0	0.0	0.0	**37.8**	35.9	25.6	0.6	
	Patient (%)		5.1	5.4	6.4	15.5	57.3	10.2	60.0

(Continues)

TABLE 3 | (Continued)

Design		Dose level 0	1	2	3	4	5	6	# of patients
Scenario 4									
	True TTB		0.57	0.83	1.09	1.35	1.6	3.16	
	True EES		0.25	0.35	0.48	0.64	0.64	0.64	
UNITED	Selection (%)	0.0	0.0	0.0	0.5	**83.0**	14.0	2.5	
	Patient (%)		5.0	5.0	6.3	55.7	20.0	7.9	60.0
gBOIN	Selection (%)	0.0	0.0	0.0	0.2	**41.0**	36.1	22.7	
	Patient (%)		5.2	5.5	6.3	8.7	18.6	55.7	60.0
Scenario 5									
	True TTB		0.57	0.83	1.09	1.35	1.60	3.16	
	True efficacy		0.02	0.08	0.14	0.24	0.37	0.37	
UNITED	Selection (%)	0.7	0.0	0.0	0.0	0.1	**91.1**	8.1	
	Patient (%)		5.1	5.1	5.1	5.3	65.7	13.7	59.7
gBOIN	Selection (%)	2.5	0.0	0.0	0.0	0.0	**56.5**	41.0	
	Patient (%)		5.2	5.5	6.3	8.7	18.6	55.7	60.0
Scenario 6									
	True TTB		0.57	0.83	1.09	1.35	1.60	2.86	
	True efficacy		0.02	0.08	0.14	0.24	0.37	0.52	
UNITED	Selection (%)	0.7	0.0	0.0	0.0	0.1	14.5	**84.7**	
	Patient (%)		5.1	5.1	5.1	5.2	14.0	65.5	59.7
gBOIN	Selection (%)	2.5	0.0	0.0	0.0	0.0	12.0	**85.5**	
	Patient (%)		5.2	5.5	6.3	8.7	16.3	58.1	60.0
Scenario 7									
	True TTB		1.49	3.04	4.54	5.17	5.79	6.41	
	True efficacy		0.32	0.48	0.49	0.52	0.59	0.76	
UNITED	Selection (%)	0.5	14.1	**82.8**	2.6	0.1	0.0	0.0	
	Patient (%)		10.6	48.7	18.3	10.9	8.1	3.4	59.8
gBOIN	Selection (%)	0.0	4.6	**90.4**	5.0	0.0	0.0	0.0	
	Patient (%)		14.2	44.3	6.8	0.3	0.0	0.0	60.0
Scenario 8									
	True TTB		0.44	1.28	3.04	4.77	5.33	5.9	
	True efficacy		0.12	0.38	0.59	0.64	0.69	0.72	
UNITED	Selection (%)	0.2	0.0	14.8	**84.9**	0.0	0.0	0.0	
	Patient (%)		5.1	11.8	59.3	11.9	8.5	3.4	59.9
gBOIN	Selection (%)	0.0	3.0	3.7	**89.7**	3.6	0.0	0.0	
	Patient (%)		5.3	12.1	75.1	7.4	0.1	0.0	60.0
Scenario 9									
	True TTB		0.44	1.28	3.04	4.78	5.33	5.9	
	True efficacy		0.05	0.1	0.2	0.36	0.48	0.7	
UNITED	Selection (%)	**99.8**	0.0	0.0	0.1	0.1	0.0	0.0	
	Patient (%)		13.8	13.8	15.3	27.6	20.7	8.8	26.8
gBOIN	Selection (%)	**99.4**	0.0	0.0	0.0	0.6	0.0	0.0	
	Patient (%)		5.3	12.1	75.2	7.4	0.1	0.0	60.0

TABLE 4 | The operating characteristics (OCs) of the UNITED and TITE-UNITED designs with delayed outcomes, including percentages of OBD selection and patients' allocation, number of patients in the trial and trial duration, based on 10,000 simulated trials.

		Dose level							# of	Duration
Design		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**patients**	**(in months)**
Scenario 1										
	True ETS^S		0.08	0.12	0.22	0.32	0.5	0.57		
	True EES^S		0.45	0.45	0.45	0.45	0.48	0.48		
TITE-UNITED	Selection (%)	6.8	**71.1**	15.2	4.4	2.4	0.1	0.0		
	Patient (%)		48.5	17.7	12.1	10.6	7.9	3.2	58.6	36.6
UNITED	Selection (%)	5.8	**71.3**	15.2	4.9	2.7	0.1	0.0		
	Patient (%)		46.9	17.6	12.6	11.4	8.2	3.2	58.1	69.0
Scenario 2										
	True ETS^S		0.05	0.14	0.21	0.33	0.42	0.53		
	True EES^S		0.30	0.50	0.50	0.51	0.51	0.51		
TITE-UNITED	Selection (%)	3.4	25.1	**64.0**	6.2	1.1	0.3	0.0		
	Patient (%)		30.2	35.1	13.1	10.0	8.1	3.5	58.9	37.0
UNITED	Selection (%)	4.0	22.4	**64.5**	6.9	1.8	0.5	0.0		
	Patient (%)		27.8	36.9	13.2	10.4	8.2	3.6	58.8	70.0
Scenario 3										
	True ETS^S		0.05	0.08	0.10	0.20	0.28	0.37		
	True EES^S		0.13	0.30	0.58	0.58	0.58	0.60		
TITE-UNITED	Selection (%)	0.3	1.1	16.0	**76.9**	4.9	0.7	0.1		
	Patient (%)		8.9	24.4	40.2	12.3	9.7	4.6	59.8	38.1
UNITED	Selection (%)	0.4	1.0	14.3	**78.2**	4.9	0.9	0.3		
	Patient (%)		8.1	22.2	42.9	12.3	9.7	4.6	59.8	70.1
Scenario 4										
	True ETS^S		0.05	0.08	0.10	0.12	0.20	0.37		
	True EES^S		0.07	0.13	0.20	0.45	0.45	0.46		
TITE-UNITED	Selection (%)	3.2	0.1	0.7	7.8	**77.7**	8.9	1.7		
	Patient (%)		6.5	11.6	21.6	39.7	14.5	6.2	59.1	40.5
UNITED	Selection (%)	3.8	0.1	0.3	4.3	**77.7**	11.5	2.3		
	Patient (%)		5.8	8.9	18.1	44.7	15.8	6.7	58.9	72.2
Scenario 5										
	True ETS^S		0.05	0.08	0.10	0.12	0.13	0.25		
	True EES^S		0.03	0.07	0.13	0.22	0.46	0.46		
TITE-UNITED	Selection (%)	3.8	0.1	0.1	0.6	11.4	**78.6**	5.5		
	Patient (%)		5.4	7.4	13.5	25.7	40.1	7.8	58.7	40.9
UNITED	Selection (%)	4.1	0.2	0.0	0.2	6.3	**81.4**	7.7		
	Patient (%)		5.4	6.2	10.3	21.8	47.3	9.0	58.7	72.4
Scenario 6										
	True ETS^S		0.05	0.08	0.10	0.12	0.13	0.18		
	True EES^S		0.03	0.07	0.13	0.18	0.23	0.48		
TITE-UNITED	Selection (%)	7.7	0.9	0.0	0.3	5.1	15.8	**70.2**		
	Patient (%)		5.8	7.2	11.9	15.8	26.4	32.9	58.1	41.3

(Continues)

TABLE 4 | (Continued)

Design		Dose level 0	1	2	3	4	5	6	# of patients	Duration (in months)
UNITED	Selection (%)	9.5	1.5	0.1	0.2	1.7	13.7	**73.2**		
	Patient (%)		5.7	6.4	9.2	15.4	25.0	38.2	57.5	71.1
Scenario 7										
	True ETS^S		0.10	0.27	0.39	0.50	0.59	0.77		
	True EES^S		0.27	0.53	0.55	0.63	0.73	0.73		
TITE-UNITED	Selection (%)	13.0	10.6	**69.3**	6.1	1.0	0.0	0.0		
	Patient (%)		28.4	47.6	12.9	6.7	3.3	1.0	56.3	32.5
UNITED	Selection (%)	14.4	9.8	**69.5**	5.5	0.8	0.0	0.0		
	Patient (%)		26.9	49.6	12.6	6.7	3.2	0.9	55.7	66.2
Scenario 8										
	True ETS^S		0.05	0.13	0.28	0.44	0.52	0.60		
	True EES^S		0.07	0.25	0.52	0.52	0.62	0.63		
TITE-UNITED	Selection (%)	16.8	0.4	10.1	**69.8**	3.2	0.4	0.0		
	Patient (%)		8.7	27.2	43.3	11.2	6.9	2.7	55.7	35.0
UNITED	Selection (%)	14.7	0.4	9.6	**73.1**	2.2	0.1	0.0		
	Patient (%)		8.2	26.0	45.6	10.9	6.7	2.6	54.8	65.6
Scenario 9										
	True ETS^S		0.22	0.30	0.40	0.50	0.53	0.60		
	True EES^S		0.10	0.13	0.22	0.30	0.42	0.53		
TITE-UNITED	Selection (%)	**93.2**	0.0	0.2	3.2	3.0	0.4	0.1		
	Patient (%)		25.0	26.7	25.3	14.0	6.8	2.2	30.1	23.9
UNITED	Selection (%)	**96.1**	0.0	0.0	1.5	1.9	0.4	0.0		
	Patient (%)		23.1	25.6	25.6	15.6	7.6	2.5	26.2	32.8

within the toxicity assessment window, we then generated the time to the second toxicity outcome as the sum of a sample from the Weibull distribution and the time to the first toxicity outcome. This process was repeated until the time to subsequent toxicity outcomes extended beyond the toxicity assessment window. We applied the same approach to time-to-efficacy outcomes. The highest observed toxicity and efficacy grades at each interim assessment were used to make dose assignment decisions in the TITE-UNITED design. It is important to note that when a patient experienced the highest observed toxicity and efficacy grades, they were considered to have a complete outcome. The simulation results, presented in Table S5, demonstrated that the TITE-UNITED design maintained its efficiency, even in situations where multiple toxicity and efficacy events occurred.

4.4 | Sensitivity Analysis

4.4.1 | Umbrella Shaped Dose-Efficacy Relationship

Our designs assume that the dose-efficacy curve increases initially and then plateaus. We performed a sensitivity analysis under settings similar to those described in Section 4.1, evaluating the performance of the UNITED design with quasi-binary toxicity and efficacy endpoints when the dose-efficacy curve was umbrella-shaped, as illustrated in Scenarios 1–5 of Table 5. Specifically, Scenarios 1–3 represented scenarios where the MTD and OBD did not match, while Scenarios 4–5 depicted cases where they did match. The UNITED design consistently outperformed the gBOIN-ET, gBOIN, and quasi-CRM designs in terms of OBD selection and patient allocation. Additionally, a similar sensitivity analysis was conducted for continuous toxicity and efficacy endpoints as described in Section 4.2. The UNITED design exhibited superior performance in scenarios where the MTD and OBD did not match (Scenarios 6–8) and slightly less effective performance when MTD and OBD were matched (Scenarios 9–10).

4.4.2 | Correlation Between Toxicity and Efficacy

In developing the probability models and methodologies for the UNITED design, we assumed independence between toxicity and efficacy outcomes. Given the use of multiple experimental doses and the assumption of a non-decreasing dose-response for these outcomes, this assumption does not necessarily imply marginal independence [38]. To address potential concerns about the association between toxicity and efficacy, we conducted a sensitivity analysis. We employed a

TABLE 5 | Sensitivity analysis of the UNITED design under umbrella-shaped dose-efficacy relationship.

Design		Dose level							# of patients
		0	**1**	**2**	**3**	**4**	**5**	**6**	
Scenario 1									
	True ETS^S		0.03	0.17	0.23	0.27	0.45	0.53	
	True EES^S		0.13	0.32	0.57	0.54	0.48	0.44	
UNITED	Selection (%)	7.0	0.5	28.7	**62.1**	1.5	0.1	0.0	
	Patient (%)		10.1	31.8	35.4	10.2	8.5	3.9	57.8
gBOIN-ET	Selection (%)	3.7	0.6	21.9	**59.8**	13.5	0.5	0.1	
	Patient (%)		7.6	31.9	49.3	10.3	0.7	0.2	59.0
gBOIN	Selection (%)	2.4	0.9	12.7	**57.7**	25.0	1.2	0.0	
	Patient (%)		5.6	14.1	25.1	37.0	16.5	1.6	60.0
quasi-CRM	Selection (%)	1.8	1.8	13.2	**56.0**	26.1	1.1	0.0	
	Patient (%)		5.2	16.7	25.3	40.7	12.0	0.0	60.0
Scenario 2									
	True ETS^S		0.03	0.11	0.13	0.17	0.25	0.33	
	True EES^S		0.13	0.23	0.40	0.67	0.63	0.57	
UNITED	Selection (%)	0.6	0.9	6.0	33.0	**58.2**	1.1	0.1	
	Patient (%)		7.6	15.1	28.7	33.3	10.4	5.0	59.8
gBOIN-ET	Selection (%)	0.2	0.0	1.5	42.8	**52.7**	2.5	0.2	
	Patient (%)		6.3	12.3	42.0	37.5	1.7	0.2	60.0
gBOIN	Selection (%)	0.2	0.4	2.3	12.0	**51.4**	29.2	4.6	
	Patient (%)		5.3	6.5	7.2	14.0	33.3	33.7	60.0
quasi-CRM	Selection (%)	0.1	0.3	2.1	12.3	**56.6**	27.7	0.9	
	Patient (%)		5.2	7.3	7.5	22.7	51.0	6.3	60.0
Scenario 3									
	True ETS^S		0.03	0.11	0.13	0.15	0.23	0.27	
	True EES^S		0.03	0.07	0.13	0.30	0.51	0.46	
UNITED	Selection (%)	3.3	0.0	0.1	0.5	31.2	**61.4**	3.5	
	Patient (%)		5.5	6.2	11.1	34.2	35.9	7.1	59.1
gBOIN-ET	Selection (%)	7.5	0.2	0.3	0.4	18.8	**58.5**	14.3	
	Patient (%)		6.2	6.8	8.0	26.5	43.0	9.4	58.4
gBOIN	Selection (%)	1.6	0.2	0.4	2.6	16.8	**58.3**	20.1	
	Patient (%)		5.3	6.5	7.1	11.7	26.4	43.0	60.0
quasi-CRM	Selection (%)	0.6	0.6	1.3	8.4	26.3	**59.3**	3.5	
	Patient (%)		5.2	7.3	7.5	19.2	52.4	8.5	60.0
Scenario 4									
	True ETS^S		0.03	0.17	0.30	0.40	0.50	0.53	
	True EES^S		0.23	0.30	0.53	0.70	0.60	0.53	
UNITED	Selection (%)	10.9	5.8	20.8	**57.8**	4.7	0.0	0.0	
	Patient (%)		16.1	26.8	35.9	12.1	6.5	2.6	57.2
gBOIN-ET	Selection (%)	4.2	10.9	30.9	**45.3**	6.5	1.4	0.6	
	Patient (%)		16.6	36.4	37.6	6.9	1.7	0.7	59.0
gBOIN	Selection (%)	5.2	8.7	10.7	**58.0**	17.3	0.2	0.0	
	Patient (%)		5.7	19.9	48.4	21.8	3.9	0.4	60.0
quasi-CRM	Selection (%)	2.8	11.0	12.4	**58.6**	15.2	0.0	0.0	
	Patient (%)		5.2	22.1	48.8	21.8	1.9	0.1	60.0

(Continues)

TABLE 5 | (Continued)

Design		Dose level 0	1	2	3	4	5	6	# of patients
Scenario 5									
	True ETS^S		0.17	0.30	0.40	0.50	0.53	0.60	
	True EES^S		0.30	0.53	0.70	0.60	0.53	0.40	
UNITED	Selection (%)	13.3	18.0	**64.1**	4.5	0.0	0.0	0.0	
	Patient (%)		30.8	43.9	12.7	6.6	4.4	1.7	56.3
gBOIN-ET	Selection (%)	9.9	38.9	**43.6**	5.6	1.4	0.5	0.1	
	Patient (%)		43.0	44.2	8.9	2.3	1.1	0.5	56.1
gBOIN	Selection (%)	5.3	11.1	**66.0**	17.3	0.2	0.0	0.0	
	Patient (%)		19.8	53.8	21.9	4.0	0.5	0.0	60.0
quasi-CRM	Selection (%)	6.8	13.5	**66.0**	13.6	0.1	0.0	0.0	
	Patient (%)		24.9	55.4	17.3	2.4	0.1	0.0	59.9
Scenario 6									
	True TTB		0.22	0.91	1.48	3.04	4.81	6.1	
	True efficacy		0.33	0.58	0.55	0.52	0.5	0.48	
UNITED	Selection (%)	0.0	0.0	**94.1**	4.9	1.0	0.0	0.0	
	Patient (%)		5.1	60.1	14.0	10.5	7.3	2.9	60.0
gBOIN	Selection (%)	0.0	0.0	**62.4**	32.3	5.4	0.0	0.0	
	Patient (%)		5.1	5.7	11.4	73.0	4.7	0.1	60.0
Scenario 7									
	True TTB		0.32	0.74	1.25	1.57	3.04	4.74	
	True efficacy		0.30	0.45	0.65	0.6	0.58	0.45	
UNITED	Selection (%)	0.0	0.0	0.1	**95.9**	3.1	0.8	0.0	
	Patient (%)		3.0	3.8	35.7	7.9	6.6	3.0	60.0
gBOIN	Selection (%)	0.0	0.0	0.2	**66.4**	26.2	7.2	0.0	
	Patient (%)		5.1	5.4	6.4	15.5	57.3	10.2	60.0
Scenario 8									
	True TTB		0.57	0.83	1.09	1.35	1.60	2.86	
	True efficacy		0.2	0.3	0.4	0.6	0.55	0.5	
UNITED	Selection (%)	0.0	0.0	0.0	0.3	**95.8**	3.5	0.4	
	Patient (%)		5.0	5.1	5.8	64.5	13.8	5.9	60.0
gBOIN	Selection (%)	0.0	0.0	0.0	0.1	**74.8**	24.0	1.1	
	Patient (%)		5.2	5.5	6.3	8.7	16.3	58.1	60.0
Scenario 9									
	True TTB		1.49	3.04	4.54	5.17	5.79	6.41	
	True efficacy		0.4	0.6	0.55	0.5	0.48	0.45	
UNITED	Selection (%)	0.1	15.6	**84.0**	0.3	0.0	0.0	0.0	
	Patient (%)		12.7	59.4	11.0	7.9	6.3	2.8	59.9
gBOIN	Selection (%)	0.0	8.6	**90.9**	0.5	0.0	0.0	0.0	
	Patient (%)		14.2	73.9	11.3	0.6	0.0	0.0	60.0
Scenario 10									
	True TTB		0.44	1.28	3.04	4.77	5.33	5.90	
	True efficacy		0.12	0.38	0.59	0.55	0.50	0.45	
UNITED	Selection (%)	0.2	0.0	15.6	**84.2**	0.0	0.0	0.0	
	Patient (%)		5.1	13.0	62.5	9.2	7.1	3.1	59.9
gBOIN	Selection (%)	0.0	0.0	9.7	**90.2**	0.1	0.0	0.0	
	Patient (%)		5.3	12.1	75.1	7.4	0.1	0.0	60.0

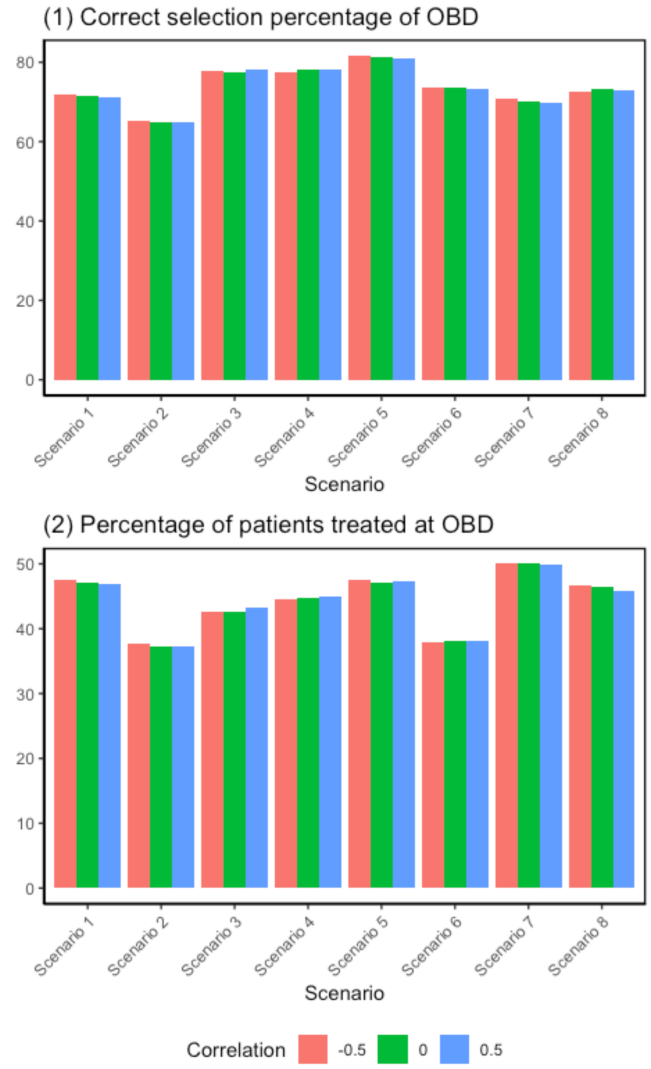

FIGURE 2 | Sensitivity analysis for the correlation between quasi-binary toxicity and efficacy endpoints.

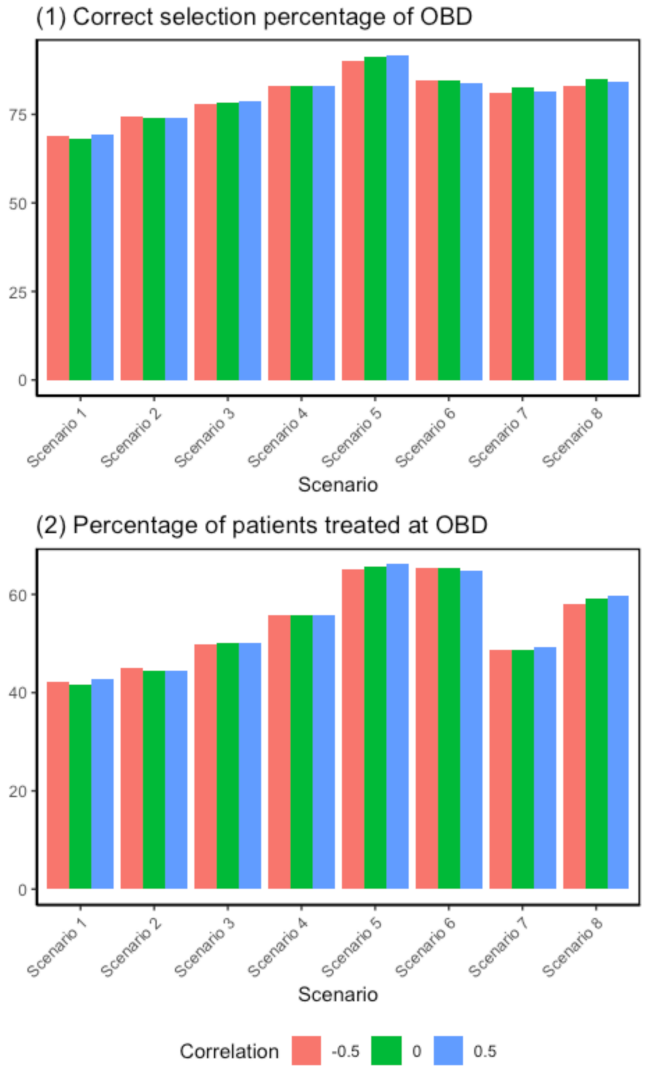

FIGURE 3 | Sensitivity analysis for the correlation between continous toxicity and efficacy endpoints.

Gaussian copula with a prespecified correlation ρ to simulate data [39]. The detailed methods for data generation are described in the Supporting Information. We conducted a sensitivity analysis to test the independence assumption between toxicity and efficacy by varying correlations from negative to positive, as illustrated in Figure 2 and Figure 3. The results of this analysis demonstrated the robustness of our independence assumption between toxicity and efficacy endpoints.

4.4.3 | Mixed Types of Responses of Toxicity and Efficacy

We conducted a sensitivity analysis for scenarios involving mixed types of endpoints, such as quasi-binary for toxicity and continuous for efficacy. The UNITED design demonstrated desirable performance across these varied settings. Detailed descriptions of the scenarios and the corresponding simulation results are summarized and discussed in the Supporting Information.

4.4.4 | Sensitivity Analysis of TITE-UNITED Design

We conducted a comprehensive sensitivity analysis of the TITE-UNITED design through simulation studies. Specifically, we explored different distributions for generating time-to-toxicity and time-to-efficacy events, including Log-logistic, Uniform, and Weibull, as shown in Figure S1. In Figure S2, we varied the patient enrollment rates, testing cases with 2, 3, and 4 patients per month. In Figure S3, we examined different lengths for the assessment windows for toxicity and efficacy, testing combinations of [1 month, 2 months], [1.5 months, 3 months], and [3 months, 3 months]. The simulation results indicate that the TITE-UNITED design is robust against variations in patient arrival distributions, enrollment rates, and time-to-event distributions. When the assessment window for efficacy significantly exceeds that for toxicity, such as an efficacy assessment window of 6 months, we conducted additional sensitivity analyses by varying the enrollment suspension rule (i.e., required percentages of completed toxicity and efficacy outcomes). The results are summarized in Table S6. We recommend modifying the enrollment suspension rule to require that at least 50% of toxicity outcomes and at least

30% of efficacy outcomes are completely observed. This adjustment substantially reduces the trial duration while only moderately affecting the OBD selection and patient allocation.

5 | Discussion

We developed the UNITED design as a comprehensive and flexible extension of the mISO design, capable of incorporating ordinal graded, continuous, and mixed toxicity and efficacy endpoints for dose-finding and optimization. Additionally, we enhanced this approach by introducing the TITE-UNITED design to accommodate delayed toxicity and efficacy outcomes using an approximated likelihood approach. The UNITED design, which makes no parametric assumptions on dose-response relationships, consistently delivers desirable performance across a range of clinically meaningful dose-response curves. Its clear, clinically interpretable model expression and dose-finding algorithm facilitate the translation from statistical methodologies to clinical applications. Simulation studies confirm that both UNITED and TITE-UNITED designs maintain robust operating characteristics in various clinical scenarios. Furthermore, sensitivity analysis shows that these designs perform well under different conditions, including various correlations between toxicity and efficacy, umbrella-shaped dose-efficacy curves, and delayed outcome scenarios.

For composite scores outcomes like TTB, specifying a targeted value requires special consideration due to the lack of a straightforward clinical interpretation for TTB. This complexity is compounded by the challenge of eliciting TTB directly from clinicians, a concern highlighted by Chen et al. [7]. Bekele and Thall [8] propose a procedure where clinicians define acceptable toxicity outcomes for multiple scenarios, from which the target TTB is derived as the mean. Although this method systematically establishes the target TTB, it involves challenges due to the complexity of considering various toxicity types and grades. Clinicians often struggle to consistently visualize multiple acceptable toxicity scenarios, resulting in significant variability in TTB. Furthermore, the composite scores are not true continuous variables. Those issues also extend to other composite scores like TTP, leading to a preference for ETS and the quasi-binomial likelihood approach in practical settings.

An interesting future direction for these designs is the consideration of optimal dynamic treatment regimes, which could adjust a patient's dose across different treatment cycles. This is particularly relevant for trials involving single-agent therapies, molecularly targeted agents, and immunotherapies, often combined with other agents. A multinomial likelihood may precisely distinguish the contributions according to different toxicity and efficacy grades. Furthermore, although sensitivity analysis demonstrates that the TITE-UNITED design performs well even with multiple observed grades throughout the trial, extending the TITE-UNITED design to develop a multinomial likelihood-based approach that incorporates multiple grade events – assuming that the toxicity and efficacy grades are not mutually exclusive – could prove to be a valuable enhancement. There is significant interest in extending the UNITED and TITE-UNITED frameworks to determine the optimal dose combination in Phase I/II drug-combination trials. In this paper, we define the OBD as efficacy-driven, following recent methodological review [40]. The potential to develop utility-based versions of UNITED and TITE-UNITED to incorporate utility considerations into dose optimization is substantial. If the OBD is defined in terms of a utility function, then correlations among delayed outcomes should be considered.

Acknowledgments

The authors thank the Editor, Associate Editor, and the Referees for their thoughtful and constructive comments and suggestions. The research of Yingjie Qiu is partially supported by NIH grant P30CA142543.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. B. E. Storer, "Design and Analysis of Phase I Clinical Trials," *Biometrics* 45, no. 3 (1989): 925–937, https://doi.org/10.2307/2531693.

2. J. O'Quigley, M. Pepe, and L. Fisher, "Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer," *Biometrics* 46, no. 1 (1990): 33–48, https://doi.org/10.2307/2531628.

3. S. Liu and Y. Yuan, "Bayesian Optimal Interval Designs for Phase I Clinical Trials," *Journal of the Royal Statistical Society: Series C: Applied Statistics* 64, no. 3 (2015): 507–523, https://doi.org/10.1111/rssc.12089.

4. J. R. Brahmer, C. G. Drake, I. Wollner, et al., "Phase I Study of Single-Agent Anti–Programmed Death-1 (MDX-1106) in Refractory Solid Tumors: Safety, Clinical Activity, Pharmacodynamics, and Immunologic Correlates," *Journal of Clinical Oncology* 28, no. 19 (2010): 3167–3175.

5. C. Le Tourneau, V. Diéras, P. Tresca, W. Cacheux, and X. Paoletti, "Current Challenges for the Early Clinical Development of Anticancer Drugs in the Era of Molecularly Targeted Agents," *Targeted Oncology* 5 (2010): 65–72.

6. N. Penel, A. Adenis, S. Clisant, and J. Bonneterre, "Nature and Subjectivity of Dose-Limiting Toxicities in Contemporary Phase 1 Trials: Comparison of Cytotoxic Versus Non-Cytotoxic Drugs," *Investigational New Drugs* 29 (2011): 1414–1419.

7. K. Chen, L. Wang, and Y. Yuan, "MT-Keyboard: A Bayesian Model-Assisted Interval Design to Account for Toxicity Grades and Types for Phase I Trials," *Statistics in Biopharmaceutical Research* 16, no. 3 (2024): 305–314.

8. B. N. Bekele and P. F. Thall, "Dose-Finding Based on Multiple Toxicities in a Soft Tissue Sarcoma Trial," *Journal of the American Statistical Association* 99, no. 465 (2004): 26–35.

9. S. Lee, D. Hershman, P. Martin, J. Leonard, and Y. Cheung, "Toxicity Burden Score: A Novel Approach to Summarize Multiple Toxic Effects," *Annals of Oncology* 23, no. 2 (2012): 537–541.

10. M. Ezzalfani, S. Zohar, R. Qin, S. J. Mandrekar, and M. C. L. Deley, "Dose-Finding Designs Using a Novel Quasi-Continuous Endpoint for Multiple Toxicities," *Statistics in Medicine* 32, no. 16 (2013): 2728–2746.

11. Z. Yuan, R. Chappell, and H. Bailey, "The Continual Reassessment Method for Multiple Toxicity Grades: A Bayesian Quasi-Likelihood Approach," *Biometrics* 63, no. 1 (2007): 173–179.

12. R. Mu, Y. Yuan, J. Xu, S. J. Mandrekar, and J. Yin, "gBOIN: A Unified Model-Assisted Phase I Trial Design Accounting for Toxicity Grades, and Binary or Continuous End Points," *Journal of the Royal Statistical Society: Series C: Applied Statistics* 68, no. 2 (2019): 289–308.

13. B. J. Druker, "Perspectives on the Development of a Molecularly Targeted Agent," *Cancer Cell* 1, no. 1 (2002): 31–36.

14. J. R. Sachs, K. Mayawala, S. Gadamsetty, S. P. Kang, and D. P. Alwis, "Optimal Dosing for Targeted Therapies in Oncology: Drug Development Cases Leading by Example," *Clinical Cancer Research* 22, no. 6 (2016): 1318–1324.

15. FDA U, "Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases. Draft Guidance for Industry" (2023), https://www.fda.gov/regulatory-information/search-fda-guidance-documents/optimizing-dosage-human-prescription-drugs-and-biological-products-treatment-oncologic-diseases.

16. P. F. Thall and J. D. Cook, "Dose-Finding Based on Efficacy–Toxicity Trade-Offs," *Biometrics* 60, no. 3 (2004): 684–693.

17. Y. Zang, J. J. Lee, and Y. Yuan, "Adaptive Designs for Identifying Optimal Biological Dose for Molecularly Targeted Agents," *Clinical Trials* 11, no. 3 (2014): 319–327.

18. S. Liu and V. E. Johnson, "A Robust Bayesian Dose-Finding Design for Phase I/II Clinical Trials," *Biostatistics* 17, no. 2 (2016): 249–263.

19. M. K. Riviere, Y. Yuan, J. H. Jourdan, F. Dubois, and S. Zohar, "Phase I/II Dose-Finding Design for Molecularly Targeted Agent: Plateau Determination Using Adaptive Randomization," *Statistical Methods in Medical Research* 27, no. 2 (2018): 466–479.

20. K. Takeda, M. Taguri, and S. Morita, "BOIN-ET: Bayesian Optimal Interval Design for Dose Finding Based on Both Efficacy and Toxicity Outcomes," *Pharmaceutical Statistics* 17, no. 4 (2018): 383–395.

21. R. Lin, Y. Zhou, F. Yan, D. Li, and Y. Yuan, "BOIN12: Bayesian Optimal Interval Phase I/II Trial Design for Utility-Based Dose Finding in Immunotherapy and Targeted Therapies," *JCO Precision Oncology* 4 (2020): 1393–1402.

22. Y. Qiu, Y. Zhao, H. Liu, S. Cao, C. Zhang, and Y. Zang, "Modified Isotonic Regression Based Phase I/II Clinical Trial Design Identifying Optimal Biological Dose," *Contemporary Clinical Trials* 127 (2023): 107139.

23. Y. Qiu and M. Li, "A Bayesian Dynamic Model-Based Adaptive Design for Oncology Dose Optimization in Phase I/II Clinical Trials," *Pharmaceutical Statistics* 24, no. 2 (2024): e2451, https://doi.org/10.1002/pst.2451.

24. B. N. Bekele and Y. Shen, "A Bayesian Approach to Jointly Modeling Toxicity and Biomarker Expression in a Phase I/II Dose-Finding Trial," *Biometrics* 61, no. 2 (2005): 343–354, https://doi.org/10.1111/j.1541-0420.2005.00314.x.

25. Y. Zhou, J. Whitehead, E. Bonvini, and J. W. Stevens, "Bayesian Decision Procedures for Binary and Continuous Bivariate Dose-Escalation Studies," *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 5, no. 2 (2006): 125–133.

26. B. Guo and Y. Yuan, "DROID: Dose-Ranging Approach to Optimizing Dose in Oncology Drug Development," *Biometrics* 79, no. 4 (2023): 2907–2919.

27. K. Takeda, S. Morita, and M. Taguri, "gBOIN-ET: The Generalized Bayesian Optimal Interval Design for Optimal Dose-Finding Accounting for Ordinal Graded Efficacy and Toxicity in Early Clinical Trials," *Biometrical Journal* 64, no. 7 (2022): 1178–1191.

28. K. Takeda, Y. Yamaguchi, M. Taguri, and S. Morita, "TITE-gBOIN-ET: Time-To-Event Generalized Bayesian Optimal Interval Design to Accelerate Dose-Finding Accounting for Ordinal Graded Efficacy and Toxicity Outcomes," *Biometrical Journal* 65, no. 7 (2023): 2200265.

29. R. E. Barlow, D. J. Bartholomew, Jr., J. M. Bremner, and H. D. Brunk, *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression*, 7th ed. (Wiley, 1972).

30. L. Duska, C. Krasner, D. O'Malley, et al., "A Phase Ib/II and Pharmacokinetic Study of EP0057 (Formerly CRLX101) in Combination With Weekly Paclitaxel in Patients With Recurrent or Persistent Epithelial Ovarian, Fallopian Tube, or Primary Peritoneal Cancer," *Gynecologic Oncology* 160, no. 3 (2021): 688–695, https://doi.org/10.1016/j.ygyno.2020.12.025.

31. Z. Chen, M. D. Krailo, S. P. Azen, and M. Tighiouart, "A Novel Toxicity Scoring System Treating Toxicity Response as a Quasi-Continuous Variable in Phase I Clinical Trials," *Contemporary Clinical Trials* 31, no. 5 (2010): 473–482.

32. G. Norman, "Likert Scales, Levels of Measurement and the "Laws" of Statistics," *Advances in Health Sciences Education* 15 (2010): 625–632.

33. W. F. Benedict, Z. Tao, C. S. Kim, et al., "Intravesical Ad-IFNαi Causes Marked Regression of Human Bladder Cancer Growing Orthotopically in Nude Mice and Overcomes Resistance to IFN-αi Protein," *Molecular Therapy* 10, no. 3 (2004): 525–532.

34. A. Hirakawa, "An Adaptive Dose-Finding Approach for Correlated Bivariate Binary and Continuous Outcomes in Phase I Oncology Trials," *Statistics in Medicine* 31, no. 6 (2012): 516–532.

35. N. Z. Shi and H. Jiang, "Maximum Likelihood Estimation of Isotonic Normal Means With Unknown Variances," *Journal of Multivariate Analysis* 64, no. 2 (1998): 183–195.

36. R. Lin and Y. Yuan, "Time-To-Event Model-Assisted Designs for Dose-Finding Trials With Delayed Toxicity," *Biostatistics* 21, no. 4 (2020): 807–824.

37. Y. K. Cheung and R. Chappell, "Sequential Designs for Phase I Clinical Trials With Late-Onset Toxicities," *Biometrics* 56, no. 4 (2000): 1177–1182.

38. P. K. Kimani, N. Stallard, and J. L. Hutton, "Dose Selection in Seamless Phase II/III Clinical Trials Based on Efficacy and Safety," *Statistics in Medicine* 28, no. 6 (2009): 917–936.

39. R. B. Nelsen, *An Introduction to Copulas* (Springer, 1999).

40. Y. Zang, B. Guo, Y. Qiu, H. Liu, M. Opyrchal, and X. Lu, "Adaptive Phase I–II Clinical Trial Designs Identifying Optimal Biological Doses for Targeted Agents and Immunotherapies," *Clinical Trials* 21, no. 3 (2024): 17407745231220661, https://doi.org/10.1177/17407745231220661.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.