

Two-Stage Design Sample Size Determination for Two Doses in Oncology Phase II Trials

MEIRUO XIANG¹, YANGRUIJUE MA, AND QIAN LI*

Abstract

In oncology therapy development, Simon’s two-stage design is commonly employed in early-phase clinical trials to assess the preliminary efficacy of a single dose, typically the maximum tolerable dose (MTD) or the maximum assessed dose (MAD). In this design, a dose may be terminated at the first stage if the anti-tumor activity is insufficient or it may proceed to the second stage for further evaluation with more subjects. To enhance the design for better benefit-risk profile dose selection and to meet the increasing needs for study designs that explore dose-response relationships, we extend Simon’s two-stage design to evaluate two doses and to include early termination for success in addition to futility. The proposed method derives decision rules and sample sizes for optimal study designs that minimize the expected or overall sample sizes while controlling type I error and meeting desired power.

KEYWORDS AND PHRASES: Decision rules, Dose-response, Optimal design, Two-stage.

1. BACKGROUND

In oncology phase I/II clinical trials, the primary objective is to identify the maximum tolerable or assessed dose (MTD/MAD) for new therapies and to evaluate their preliminary efficacy. These trials often employ dose escalation/de-escalation algorithms, using dose-limiting toxicity events and available safety and pharmacokinetic information to identify tolerable and therapeutically effective doses. Various methods are available for the optimal dose search, including the Bayesian optimal interval (BOIN) design for Phase I clinical trials [8], and the modified toxicity probability interval (mTPI-2) method [5]. However, these methods primarily focus on safety and may not be able to adequately evaluate the therapeutic efficacy due to the constrained sample sizes at each dose level. Typically, trials employing these designs proceed to the dose-expansion phase once the MTD/MAD dose is identified.

In trials where only one dose is identified for expansion and further efficacy assessment in phase II, Simon’s two-stage design [10] is frequently used. In this design, a dose may be terminated at the first stage if the anti-tumor activity is insufficient or it may proceed to the second stage for further evaluation with more subjects. However, although the identified dose level, often the MTD/MAD, exhibits anti-tumor activities, it may not optimize the benefit-risk ratio, which may hinder successful development in the later phase. Hence, finding a dose level that optimizes the balance

between therapeutic benefits and safety risks in the early phases is paramount for the overall success of therapy development. In recognition of this importance, the FDA issued guidance on optimizing the dosage for developing oncology therapy [12]. In the guidance “Expansion Cohorts: Use in First-in-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics Guidance for Industry” [11], the FDA advises developing statistical analysis plans that encompass not only the justification of the maximum sample size but also stopping rules for lack of activity to minimize subject exposure to ineffective treatments. Furthermore, the FDA’s guidance also encouraged exploring randomized designs for multiple doses/regimens, including justification of sample size in the early phases of therapy development.

In response to the increasing need for new statistical methodologies that explore dose-response relationships in early-phase oncology therapy development, we propose a design method that extends Simon’s single-arm two-stage design. Many modifications of Simon’s design exist in the literature. For example, Chen expanded the two-stage framework into three stages, reducing the expected sample size when the treatment is ineffective [3], while Lin and Shih introduced adaptive extensions that allow modifications based on interim results, improving flexibility in decision-making [7]. Whitehead adapted Simon’s methodology for trials with survival endpoints, broadening its applicability to time-to-event outcomes [13]. Other methods also explored balancing groups to optimize trial efficiency and minimize bias in two-stage designs. For instance, Ye and Shyr propose balanced two-stage designs that maintain equal allocation across groups, improving trial efficiency in oncology Phase II studies [14]. Parashar introduce stratified designs that

*Corresponding author.

¹The author is currently employed by the US Food and Drug Administration (FDA). Disclaimer: This article reflects the views of the authors and should not be construed to represent FDA’s views or policies.

extend Simon’s two-stage methodology to stratified patient groups, optimizing the assessment of heterogeneity across subgroups [9]. Methods were also developed to enhance Simon’s design for specific trial contexts, such as incorporating sequential monitoring and decision-making processes. Lee proposed sequential dose-finding designs that optimize patient safety by continuously evaluating efficacy and toxicity [6]. Bartroff introduced group-sequential methodologies that integrate dose-escalation strategies with efficacy evaluations, enhancing the applicability of two-stage designs in dose-response contexts [1]. Further advancements include Bayesian methods and meta-analytic methods. For example, Zohar and Chevret presented a Bayesian two-stage design for dose-ranging trials, refining dose selection through adaptive decision rules [15]. Crippa introduced meta-analytic methods for synthesizing dose-response data, which can be integrated into sequential trial designs for more comprehensive analysis [4].

Despite their advancements, existing methods are limited to evaluating a single dose and do not explore the dose-response relationships of multiple doses. In contrast, we propose to extend Simon’s design to evaluate two doses and derive decision rules to identify better doses between two doses with optimal sample sizes. The decision rules allow early termination of the study at Stage I if none of the doses show sufficient anti-tumor potential or if at least one dose exhibits optimal anti-tumor activities. The decision rules also allow the selection of one dose for further evaluation in the second stage if the anti-tumor potential is promising. The paper is organized into the following sections: an overview of Simon’s two-stage design is provided in Section 2; the new method is introduced in Section 3 along with the calculation of the decision probabilities; Section 4 presents an enumeration algorithm for sample size calculation leading to the optimal and minimax design under the constraints of overall type I error and power; and examples of the application of the new method are illustrated in Section 5.

2. OVERVIEW OF SIMON’S TWO-STAGE DESIGN

In Simon’s two-stage design, the null and alternative hypotheses are specified as $H_0 : \theta \leq \theta_0$, $H_1 : \theta \geq \theta_A$, where θ represents a response rate for anti-tumor activities such as objective response rates (ORR). The null hypothesis implies that response rates equal to θ_0 or lower are considered ineffective whereas the alternative corresponds to the parameter space that has treatment effect. A study design may consider powering certain desired response rates $\theta \geq \theta_A$. Stage I of Simon’s design uses a small sample size to explore if there is any desired anti-tumor activity. The study may be terminated at Stage I if the activity is lower than expected. Otherwise, additional subjects will be enrolled in Stage II to further evaluate the anti-tumor activity. Denote the observed number of responses at Stage I and Stage II as S_1

and S_2 , respectively, and let $S = S_1 + S_2$. The response thresholds for Stage I and Stage II are denoted as a and r , respectively, out of the sample size n_1 and n_2 , the number of subjects enrolled at Stages I and II, respectively. At the end of the first stage, if $S_1 \leq a$, the therapy is considered ineffective and the study will be terminated in Stage I. Otherwise, the study enters into Stage II. If $S > r$ at the end of the second stage, the therapy is considered effective.

Simon’s design determines the efficient sample sizes n_1 and n_2 for the decision rule a and r by controlling type I error, α , and type II error, β , at the desired levels. Let $n = n_1 + n_2$ be the total number of subjects. The expected sample size is expressed as $EN = n_1 + (1 - \text{PET})n_2$, where PET is the probability of early termination. The optimal design is derived by enumerating n , n_1 , a , r with exact binomial probabilities based on the given θ_0 , θ_A , α , and β . While controlling the type I error and achieving the desired power, Simon introduced two designs for finding sample sizes: the optimal design, which minimizes the expected sample size EN, and the minimax design, which minimizes the maximum sample size n .

3. PROPOSED METHOD

3.1 Study Design Considerations

In Simon’s two-stage design, the trial for the selected MTD/MAD dose is terminated at the first stage if the dose has low anti-tumor activities and moved to the second stage for further efficacy evaluation if adequate responses are observed in the first stage. However, recent advancements in anti-cancer treatments, such as cell and immuno-oncology therapies, show that the ORRs can be up to 80% and 90% [2]. With such high anti-tumor responses, it is reasonable to incorporate early termination for efficacy in the study design. Therefore, in addition to early termination for futility, we also propose termination for efficacy at the end of Stage I if the dose successfully demonstrates a high response rate in Stage I. By allowing the trial to be terminated for both futility and efficacy in the first stage, the proposed method could lead to additional sample size reduction in comparison to Simon’s approach. Furthermore, we extend Simon’s two-stage design to include two doses in the proposed method. The two doses will be randomized in a 1:1 ratio in the first stage assuming no prior knowledge of the preference of the two doses. Dose-response decisions will be made at the end of the first stage such that if the trial is not terminated for either futility or efficacy, only one dose will be selected to move to the next stage. In cases where the two doses show similar benefits and risk profiles, the lower dose is recommended for the next stage. The second stage of the study has a single arm.

Let θ_1 and θ_2 denote the anti-tumor responses, such as ORR, for Doses 1 and 2, respectively. The null hypothesis in the parameter space, θ_1 and θ_2 , implies that neither of the two doses is efficacious and can be written as $H_0 : \theta_1 \leq \theta_0$

and $\theta_2 \leq \theta_0$, where θ_0 represents the threshold of a sub-therapeutic anti-tumor response region. The corresponding alternative is $H_A : \theta_1 \geq \theta_A$ or $\theta_2 \geq \theta_A$. In the alternative space of θ_1 and θ_2 , we consider powering the study at the region $\theta_1 \geq \theta_A$ or $\theta_2 \geq \theta_A$, where θ_A is the expected or desired anti-tumor activity. This alternative region represents the scenario in which both doses are efficacious. The alternative space also covers the scenarios where only one dose is efficacious, such as the regions $H_{01} : \theta_1 \leq \theta_0$ vs $H_{A1} : \theta_1 \geq \theta_A$ for Dose 1 and $H_{02} : \theta_2 \leq \theta_0$ vs $H_{A2} : \theta_2 \geq \theta_A$ for Dose 2. In the proposed method, we may consider powering the study in different alternative regions. For example, a study design may power only the alternative region where $\theta_1 \geq \theta_A$ and $\theta_2 \geq \theta_A$. A more comprehensive design is to power the region of $\theta_1 \geq \theta_A$ or $\theta_2 \geq \theta_A$, which also includes the region $\theta_1 \geq \theta_A$ and $\theta_2 \geq \theta_A$. We will demonstrate in later sections that the optimal design to power the different regions will require different sample sizes.

3.2 Decision Rules

In the proposed method, we determine the decision rules and sample sizes based on the desired power and type I error rate. The decision rules include thresholds for observed responses to determine whether the trial should be terminated at the first stage for futility or success, which dose should be moved to the next stage, and whether a dose can be claimed for success at the end of Stage II. The optimal decision rule will lead to designs that use the minimum sample size to achieve the desired power for efficacious doses and control the type I error rate for the dose that lacks anti-tumor activities. Consistent with Simon's two-stage design, we refer to the design with the minimum expected sample size as the optimal design and the design with the smallest total sample size as the minimax design.

Denoting the number of patients enrolled for each dose at the first and second stage as n_1 and n_2 , respectively,

the maximum total sample size should be $n = 2n_1 + n_2$ as only one dose will be moved to the second stage. The sample size for the dose moved to the second stage is $m = n_1 + n_2$. Let S_{1k} denote the number of responses at Stage I for dose k , $k = 1$ or 2 , and S_2 denote the number of responses at Stage II for the dose that is moved to Stage II. Note in this proposed method, we only select one dose to move to Stage II. Let a_1 and r_1 be the lower and upper thresholds for terminating the study at Stage I. Doses with responses lower than or equal to a_1 will be terminated for futility, and doses with responses higher than or equal to r_1 are considered promising and will skip Stage II, moving directly to the next phase of development. Let s denote the number of cumulative responses: $S = S_{1k} + S_2$ and r is the upper threshold out of $n_1 + n_2$ for claiming efficacy for the dose at Stage II. We also assume that Dose 1 is the lower dose.

The decision rule is represented by the set of value $\{a_1, r_1, r\}$. During the conduct of the trial, the following decisions may be made based on the observed S_{1k} , and S_2 , see Figure 1 and Figure 2:

- $D1.1$ to claim either dose to be efficacious and terminate the trial at Stage I, that is, $S_{11} \geq r_1$ or $S_{12} \geq r_1$.
- $D1.2$ to drop both doses for futility and terminate the trial at Stage I, that is $S_{11} \leq a_1$ and $S_{12} \leq a_1$.
- $D2.1$ to claim Dose 1 to be efficacious at Stage II. That is $a_1 < S_{11} < r_1$ and $S_{12} \leq S_{11}$ and $S = S_{11} + S_2 > r$.
- $D2.2$ to claim Dose 2 to be efficacious at Stage II. That is $a_1 < S_{12} < r_1$ and $S_{11} < S_{12}$ and $S = S_{12} + S_2 > r$.
- $D2.3$ to claim no dose to be efficacious at Stage II. That is $S = S_{1k} + S_2 \leq r$, $a_1 < S_{1k} < r_1$, and $S_{1,3-k} \leq S_{1k}$, where $k = 1, 2$.

Assume that $S_{1k} \sim \text{Binomial}(\theta_k, n_1)$, $k = 1, 2$, follow binomial distributions and S_{11} and S_{12} are independent. We denote $\text{Bin}(s_{1k}|n_1, \theta_k) = P(S_{1k} \leq s_{1k}|n_1, \theta_k)$ as the cumu-

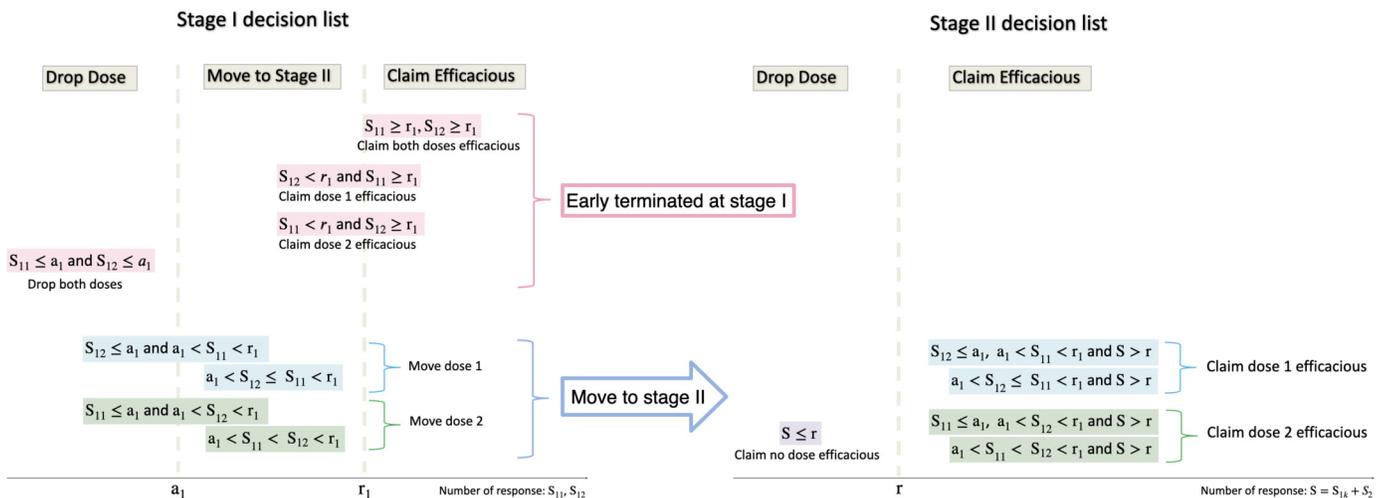


Figure 1: Decision list of the new method.

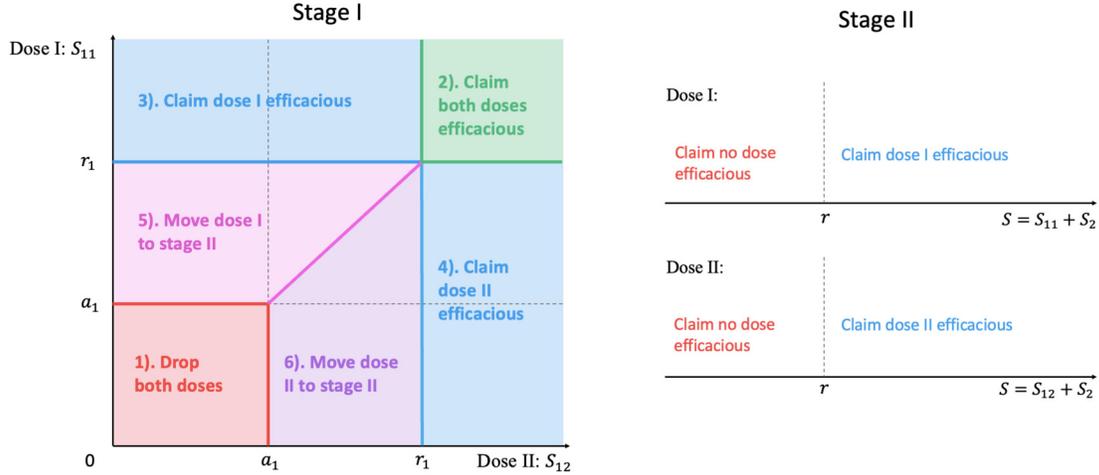


Figure 2: Two-dimensional Decision Rules. At Stage I, the x-axis represents the dose response for dose I (S_{12}) and the y-axis represents the dose response for dose II (S_{11}). According to the decision rules, the 2-dimensional space are divided into 6 parts: (1) Drop both doses for futility and terminate the trial at Stage I, for $S_{11} \leq a_1$ and $S_{12} \leq a_1$. (2) Claim both dose to be efficacious and terminate the trial at Stage I, for $S_{11} \geq r_1$ and $S_{12} \geq r_1$. (3) Claim Dose 1 to be efficacious and terminate the trial at Stage I, for $S_{11} \geq r_1$ and $S_{12} < r_1$. (4) Claim Dose 2 to be efficacious and terminate the trial at Stage I, for $S_{12} \geq r_1$ and $S_{11} < r_1$. (5) Move Dose 1 to Stage II, for $a_1 < S_{11} < r_1$ and $S_{11} \geq S_{12}$. (6) Move Dose 2 to Stage II, for $a_1 < S_{12} < r_1$ and $S_{11} < S_{12}$. If Dose 1 is moved to the Stage II, claim no dose efficacious for $S = S_{11} + S_2 \leq r$, and claim Dose 1 efficacious for $S = S_{11} + S_2 > r$. Similarly, if Dose 2 is moved to the Stage II, claim no dose efficacious for $S = S_{12} + S_2 \leq r$, and claim Dose I efficacious for $S = S_{12} + S_2 > r$.

lative probability function and $\text{bin}(s_{1k}|n_1, \theta_k) = P(S_{1k} = s_{1k}|n_1, \theta_k)$ as the probability mass function of the binomial distribution $\text{Binomial}(\theta_k, n_1)$. The probability of claiming either dose at Stage I using the decision rule $D1.1$ can be written as

$$\begin{aligned} R_1(\theta_1, \theta_2) &= P(S_{11} \geq r_1 \cup S_{12} \geq r_1 | \theta_1, \theta_2) \\ &= 1 - P(S_{11} < r_1 \cap S_{12} < r_1 | \theta_1, \theta_2) \\ &= 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \text{Bin}(r_1 - 1 | \theta_2, n_1). \end{aligned}$$

The probability of terminating the trial at Stage I for decision $D1.2$ can be written as

$$\begin{aligned} A_1(\theta_1, \theta_2) &= P(S_{11} \leq a_1 \cap S_{12} \leq a_1 | \theta_1, \theta_2) \\ &= \text{Bin}(a_1 | \theta_1, n_1) \text{Bin}(a_1 | \theta_2, n_1). \end{aligned}$$

The probability of claiming Dose 1 at Stage II for decision $D2.1$ is

$$\begin{aligned} R_{21}(\theta_1, \theta_2) &= P(a_1 < S_{11} < r_1 \cap S_{11} \geq S_{12} \cap S \geq r | \theta_1, \theta_2) \\ &= \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \text{Bin}(S_{11} | \theta_2, n_1) \\ &\quad [1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)]. \end{aligned}$$

The probability of claiming Dose 2 at Stage II for decision $D2.2$ is

$$\begin{aligned} R_{22}(\theta_1, \theta_2) &= P(a_1 < S_{12} < r_1 \cap S_{11} < S_{12} \cap S \geq r | \theta_1, \theta_2) \\ &= \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \text{Bin}(S_{12} - 1 | \theta_1, n_1) \\ &\quad [1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)]. \end{aligned}$$

Note that when $S_{11} = S_{12}$, the lower dose, which is Dose 1, will be selected to move forward.

3.3 Type I Error

The type I error rate will be discussed under the global null hypothesis denoted as $H_0 : \theta_1 \leq \theta_0 \cap \theta_2 \leq \theta_0$, which implies that none of the two doses is efficacious. The corresponding global alternative that we would like to power is $H_A : \theta_1 \geq \theta_A \cup \theta_2 \geq \theta_A$. Although the space between θ_0 to θ_A is considered alternative, we are not interested in power such alternative space. In addition to the global hypotheses, the hypotheses for the individual doses are $H_{01} : \theta_1 \leq \theta_0$ vs. $H_{A1} : \theta_1 \geq \theta_A$ and $H_{02} : \theta_2 \leq \theta_0$ vs. $H_{A2} : \theta_2 \geq \theta_A$ for doses 1 and 2, respectively. All hypotheses are 1-sided hypotheses. Under the null hypothesis $H_0 : \theta_1 \leq \theta_0$ and $\theta_2 \leq \theta_0$, the overall type I error is the maximum error rate in the null space $0 \leq \theta_1 \leq \theta_0$ and $0 \leq \theta_2 \leq \theta_0$, i.e., the maximum probability of claiming either doses or both to be active in

anti-tumor activities at Stages 1 or 2 based on the decision rules $D1.1$, $D2.1$, and $D2.2$. The need to identify the maximum type I error is because of the non-monotonic nature of the type I error function in the null space. Controlling the type I error at the level of α_0 for 1-sided tests, it can be written as formula 3.1. The derivation of the formula can be found in the appendix.

$$\begin{aligned}
& \max(R_1 + R_{21} + R_{22} | \theta_1 \leq \theta_0 \text{ and } \theta_2 \leq \theta_0) \\
& = \max\{1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
& + \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \text{Bin}(S_{11} | \theta_2, n_1) \\
& [1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
& + \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \text{Bin}(S_{12} - 1 | \theta_1, n_1) \\
& [1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)]\} \\
& \leq \alpha_0. \tag{3.1}
\end{aligned}$$

The probability of claiming Dose 1 to be efficacious under the individual null hypotheses can be written as

$$\begin{aligned}
& P_{H_{01}}(S_{11} \geq r_1) \\
& + P_{H_{01}}(a_1 < S_{11} < r_1 \cap S_{11} \geq S_{12} \cap S \geq r).
\end{aligned}$$

As only one dose can be selected to move forward, the chance of Dose 1 being the selected dose can be maximized by letting $\theta_2 = 0$. That is, $\text{Bin}(S_{11} | \theta_2, n_1) = 1$. Thus the maximum individual type I error for Dose 1 can be written as

$$\begin{aligned}
& \max[P_{H_{01}}(S_{11} \geq r_1) \\
& + P_{H_{01}}(a_1 < S_{11} < r_1 \cap S_{11} \geq S_{12} \cap S \geq r) | \\
& \theta_1 \leq \theta_0 \text{ and } \theta_2 = 0] \\
& < 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \\
& + \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \\
& [1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
& \leq \alpha_0.
\end{aligned}$$

Similarly, the probability of claiming Dose 2 is under the individual null hypotheses can be written as

$$\begin{aligned}
& P_{H_{02}}(S_{12} \geq r_1) \\
& + P_{H_{02}}(a_1 < S_{12} < r_1 \cap S_{12} > S_{11} \cap S \geq r).
\end{aligned}$$

and Dose 2 will have the maximum chance to move to Stage II when $\theta_1 = 0$. That is, $\text{Bin}(S_{12} | \theta_1, n_1) = 1$. The maximum individual type I error for Dose 2 is

$$\begin{aligned}
& \max[P_{H_{02}}(S_{12} \geq r_1) \\
& + P_{H_{02}}(a_1 < S_{12} < r_1 \cap S_{12} > S_{11} \cap S \geq r)]
\end{aligned}$$

$$\begin{aligned}
& \theta_1 = 0 \text{ and } \theta_2 \leq \theta_0 \\
& < 1 - \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
& + \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \\
& [1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)] \\
& \leq \alpha_0.
\end{aligned}$$

Note when controlling the maximum type I error for each dose, the type I error is strongly controlled.

3.4 Power

The overall power for Dose 1 and Dose 2 is calculated as the following:

$$1 - \beta = R_1 + R_{21} + R_{22} | \theta_1 = \theta_A \text{ or } \theta_2 = \theta_A.$$

In the space of $\theta_1 = \theta_A \cup \theta_2 = \theta_A$, the following three points in this alternative space will be chosen as the criteria of power: $(\theta_1 = \theta_A, \theta_2 = \theta_0)$ for the worst power of Dose 1, $(\theta_1 = \theta_0, \theta_2 = \theta_A)$ for the worst power of Dose 2, and $(\theta_1 = \theta_A, \theta_2 = \theta_A)$.

When $(\theta_1 = \theta_A, \theta_2 = \theta_0)$, the power for Dose 1 reflects the probability of selecting Dose 1 while rejecting the null hypothesis for θ_1 :

$$\begin{aligned}
& 1 - \beta_1 \\
& = P_{H_{01}}(S_{11} \geq r_1) \\
& + P_{H_{01}}(a_1 < S_{11} < r_1 \cap S_{11} \geq S_{12} \cap S \geq r) \\
& | \theta_1 = \theta_A, \theta_2 = \theta_0.
\end{aligned}$$

When $(\theta_1 = \theta_0, \theta_2 = \theta_A)$, the power for Dose 2 reflects the probability of selecting Dose 2 while rejecting the null hypothesis for θ_2 :

$$\begin{aligned}
& 1 - \beta_2 \\
& = P_{H_{02}}(S_{12} \geq r_1) \\
& + P_{H_{02}}(a_1 < S_{12} < r_1 \cap S_{12} > S_{11} \cap S \geq r) \\
& | \theta_1 = \theta_0, \theta_2 = \theta_A.
\end{aligned}$$

When $(\theta_1 = \theta_A, \theta_2 = \theta_A)$, the power reflects the combined probability of rejecting the null hypothesis for both doses:

$$\begin{aligned}
& 1 - \beta_0 \\
& = R_1 + R_{21} + R_{22} | \theta_1 = \theta_A, \theta_2 = \theta_A.
\end{aligned}$$

Detailed derivation of power can be found in the appendix.

3.5 Probability of Early Termination (PET)

The probability of early termination is the probability that the trial terminated at the first stage, by either claiming

at least one dose to be efficacious or dropping both doses for futility. PET can be calculated by

$$\begin{aligned} \text{PET} &= R_1 + A_1 \\ &= 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \text{Bin}(r_1 - 1 | \theta_2, n_1) \\ &\quad + \text{Bin}(a_1 - 1 | \theta_1, n_1) \text{Bin}(a_1 - 1 | \theta_2, n_1). \end{aligned}$$

The derivation of the formula can be found in Appendix C, where PET can be calculated under null, denoted as PET_1 , and alternative $\theta_1 = \theta_A$ and $\theta_2 = \theta_A$, denoted as PET_2 .

4. DERIVATION OF THE OPTIMAL DESIGN

To control the maximum type I error in the null space $0 \leq \theta_1 \leq \theta_0$ and $0 \leq \theta_2 \leq \theta_0$, it is necessary to search the entire null space as the type I error function in 3.1 is not monotone. To enhance the searching speed for the maximum type I error, we divided the search algorithm into two parts. The first part grids the partial sample size and decision rule, n_1 and r_1 , in the null space $0 \leq \theta_1 \leq \theta_0$ and $0 \leq \theta_2 \leq \theta_0$ by an interval of 0.02. The partial type I error, $1 - \text{Bin}(r_1 | \theta_1, n_1) \text{Bin}(r_1 | \theta_2, n_1)$, is calculated for n_1 and r_1 . Note that n_1 is the sample size of each dose during the first stage, which is usually small, therefore, n_1 is limited to less than 50 subjects. The choice of r_1 is bound to be less or equal to n_1 . In addition, only the combinations of n_1 and r_1 that control the maximum partial type I error in the null space are kept for the next part of search that completes the derivation of the full decision rules and sample size. The first step eliminates a significant portion of the n_1 and r_1 combinations and therefore enhances the efficiency of the enumeration process to identify the decision rules in the second step.

The second step of the algorithm iterates over the selected set of (n_1, r_1) from the first step to identify the combinations

of (n_2, a_1, r) that satisfy the constraints: maximum type I error less than α and power greater than $1 - \beta$. To ensure the maximum type I error is less than α for each set of decision rules and sample sizes (n_1, r_1, n_2, a_1, r) , type I errors are calculated over the null space: $\theta_{10} \in (0, \theta_0)$ and $\theta_{20} \in (0, \theta_0)$ by a step of 0.01. Decision rules with the overall type I error greater than α are eliminated. The selected decision rules require the control of the maximum individual type I errors to be less than α for each individual dose. As a result, the type I errors are strongly controlled.

Designs are not considered adequate if a_1 and r_1 are too close to each other, for example, only 1 response apart. Such designs terminate the study at Stage I either for futility or efficacy without leave much room for the grey zone (the response is between a_1 and r_1), which would warrants further evaluation in Stage II. To ensure that designs don't make a black-or-white decision but allow the dose to move on to Stage II when responses are in between the two termination criteria, additional rules, such as $r_1 > a_1 + 2$, are imposed in design selections. In addition, when the sample size assigned to the first stage n_1 is too small relative to the total number of sample size n , such designs will not be adequate for dose-response evaluation in the first stage. Therefore, further restrictions, such as $2n_1 \geq n_2 \geq n_1/2$, are also placed in the derivation algorithm. For the decision rule in Stage II, r , it is natural to require $r > r_1$.

Each study design will include sample size n_1 and n as well as decision rules a_1, r_1, r and satisfy both the type I error and power constraints. The expected sample size EN can be calculated as $\text{EN} = 2n_1 + (1 - \text{PET})n_2$. Among all decision rules that satisfy the constraints, the decision rule that yields the minimum total sample size will be selected as the minimax design and the decision rule that yields the minimum EN is the optimal design. Figure 3 illustrates the

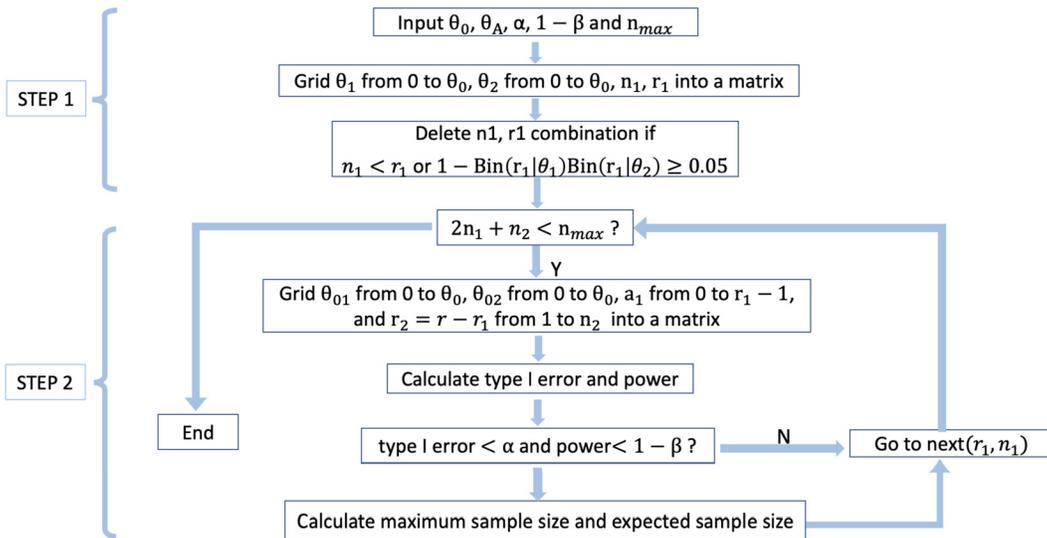


Figure 3: Algorithm of Decision Rules.

algorithm of decision rules. Here, EN can be calculated under null or alternative, denoted as EN_1 and EN_2 , respectively. In addition, an average EN_{Avg} can also be calculated as the average of EN_1 and EN_2 .

5. RESULTS

This section illustrates the results using the proposed algorithm and the selection of efficient designs that are considered optimal. Various design parameters θ_0 and θ_A are used as examples. The type I error is controlled at 1-sided 0.05 and the required power is 80%. For the optimal design, the average expected sample size EN_{Avg} is used to select the design. The optimal designs are selected from all designs that satisfy the constraints of Type I error and the required power. The corresponding design selected from the algorithm for each set of design parameters are listed in Appendix D.

Tables 1 shows the results of the optimal and minimax design when powering at $\theta_1 = \theta_A$ and $\theta_2 = \theta_A$, the same targeting anti-tumor activities for both doses, with relative treatment differences $\theta_A - \theta_0$ of either 0.3 or 0.2. Table 1 includes the total sample size, n , the sample size of the first stage n_1 , the sample size of the second stage n_2 , the termination rules a_1 and r_1 for responses, as well as the total responses for claiming efficacy from Stage II, r , the probability of early termination PET_{Avg} at Stage I, and the expected sample size EN_{Avg} . The study designs presented in Table 1 power the alternative points at the minimum of 80% and control the type I error at the level of 1-sided 0.05.

Notice that in Table 1, when $\theta_A - \theta_0 = 0.3$, the best design with a minimal total sample size, the minimax design, and the minimal expected sample size, the optimal design, for most design scenarios are the same. For the scenario where $\theta_A = 0.5$ and $\theta_0 = 0.2$, the first stage consists of $n_1 = 6$ subjects for both minimax and optimal design. If responses

are no more than 1 ($a_1 = 1$) in both doses, indicating that the compound in both doses does not exhibit sufficient anti-tumor activity, then the trial can be terminated for futility. On the other hand, if there are more than 4 ($r_1 = 4$) responses in one or both doses, the corresponding doses are considered efficacious and the trial can be terminated after the first stage for efficacy. Otherwise, if the response is in between the two thresholds a_1 and r_1 , the dose with a better response rate or the lower dose when both doses have the same response rate is moved to the second stage, and the trial continues to accrual more subjects up to $n_2 = 8$ in Stage II. The dose is considered efficacious at the end of Stage II if there are total $r = 6$ responses from both stages out of $n_1 + n_2 = 14$ subjects. With this design, the total sample size is $2n_1 + n_2 = 20$ and the average sample size is $EN_{Avg} = 16$ with $PET_{Avg} = 0.52$.

It is also observed in Table 1 that as expected, when the response differences $\theta_A - \theta_0$ become smaller, larger total and expected sample sizes are needed to meet the desired power and type I error. In design scenarios with a low threshold of θ_0 , for example, $\theta_0 = 0.2$, the study will terminate for failure in the first stage with relatively lower a_1 and r_1 as opposed to larger θ_0 . Such design makes sense as lower θ_0 may indicate diseases that have no good treatments, therefore a treatment with a lower response rate r_1 will be acceptable and the study will be terminated for futility if the treatment response is very low (with a lower threshold a_1).

Table 2 shows the results of the minimax and optimal design when powering the regions $\theta_1 = \theta_A$ or $\theta_2 = \theta_A$. In those designs, the studies are powered when either dose or both doses have response rates θ_A . As a result of this powering strategy, the overall sample size has increased by close to 50% in comparison to the design that powers the alternative $\theta_1 = \theta_A$ and $\theta_2 = \theta_A$. Similar to Table 1 for treatment difference $\theta_A - \theta_0 = 0.3$, the optimal and minimax designs are the same design.

Table 1. Power the Region of $\theta_1 \geq \theta_A$ and $\theta_2 \geq \theta_A$.

Difference	Parameter	Method	n	n_1	n_2	a_1	r_1	r	PET_1	PET_2	PET_{Avg}	EN_{Avg}	
0.3	$\theta_0 = 0.2, \theta_A = 0.5$	Minimax	20	6	8	1	4	7	0.46	0.58	0.52	16	
		Optimal	20	6	8	1	4	7	0.46	0.58	0.52	16	
	$\theta_0 = 0.3, \theta_A = 0.6$	Minimax	21	7	7	3	6	8	0.77	0.38	0.57	17	
		Optimal	21	7	7	3	6	8	0.77	0.38	0.57	17	
	$\theta_0 = 0.4, \theta_A = 0.7$	Minimax	24	7	10	3	6	12	0.54	0.57	0.55	19	
		Optimal	24	7	10	3	6	12	0.54	0.57	0.55	19	
	$\theta_0 = 0.5, \theta_A = 0.8$	Minimax	22	7	8	4	7	12	0.61	0.40	0.51	18	
		Optimal	22	7	8	4	7	12	0.61	0.40	0.51	18	
	0.2	$\theta_0 = 0.2, \theta_A = 0.4$	Minimax	41	11	19	3	6	11	0.73	0.52	0.62	30
			Optimal	41	11	19	3	6	11	0.73	0.52	0.62	30
$\theta_0 = 0.3, \theta_A = 0.5$		Minimax	52	20	12	8	11	16	0.82	0.72	0.77	43	
		Optimal	54	14	26	5	9	19	0.63	0.42	0.52	29	
$\theta_0 = 0.4, \theta_A = 0.6$		Minimax	55	21	13	11	14	20	0.86	0.67	0.77	46	
		Optimal	59	15	29	8	11	24	0.84	0.54	0.69	40	
$\theta_0 = 0.5, \theta_A = 0.7$		Minimax	53	19	15	12	15	23	0.84	0.11	0.48	46	
		Optimal	56	15	26	9	12	28	0.76	0.58	0.67	39	

Table 2. Power the Region of $\theta_1 \geq \theta_A$ or $\theta_2 \geq \theta_A$.

Difference	Parameter	Method	n	n_1	n_2	a_1	r_1	r	PET_1	PET_2	PET_{Avg}	EN_{Avg}
0.3	$\theta_0 = 0.2, \theta_A = 0.5$	Minimax	37	10	17	2	6	11	0.47	0.61	0.54	28
		Optimal	37	10	17	2	6	11	0.47	0.61	0.54	28
	$\theta_0 = 0.3, \theta_A = 0.6$	Minimax	39	12	15	4	8	14	0.54	0.69	0.62	30
		Optimal	39	12	15	4	8	14	0.54	0.69	0.62	30
	$\theta_0 = 0.4, \theta_A = 0.7$	Minimax	42	11	20	5	9	19	0.58	0.53	0.56	31
		Optimal	42	11	20	5	9	19	0.58	0.53	0.56	31
$\theta_0 = 0.5, \theta_A = 0.8$	Minimax	37	11	15	6	10	19	0.54	0.54	0.54	29	
	Optimal	37	11	15	6	10	19	0.54	0.54	0.54	29	
0.2	$\theta_0 = 0.2, \theta_A = 0.4$	Minimax	77	25	27	6	10	18	0.64	0.83	0.73	57
		Optimal	77	25	27	6	10	18	0.64	0.83	0.73	57
	$\theta_0 = 0.3, \theta_A = 0.5$	Minimax	92	27	38	9	14	29	0.56	0.75	0.66	68
		Optimal	92	27	38	9	14	29	0.56	0.75	0.66	68
	$\theta_0 = 0.4, \theta_A = 0.6$	Minimax	97	38	21	18	22	33	0.78	0.90	0.84	80
		Optimal	98	32	34	15	20	36	0.71	0.72	0.74	74
	$\theta_0 = 0.5, \theta_A = 0.7$	Minimax	94	33	28	19	23	40	0.76	0.85	0.80	72
		Optimal	95	27	41	15	20	44	0.63	0.66	0.64	69

It is interesting to observe from both Tables 1 and 2 that even though the maximum total sample sizes are similar in both optimal and the minimax designs, the expected sample sizes and the decision rules can be very different between the two. In the optimal design, the sample size allocated in the first stage is smaller than the minimax design, and a_1 , the threshold to terminate the study to accept the null, is lower than that of the minimax design in general. Because of the lower sample size in the first stage, the threshold of terminating the study for efficacy, r_1 , is also lower than that of the minimax design. Overall, when the two designs are different, the minimax designs invest more subjects in the first stage and has a higher probability to terminate under null and some cases under alternative, relative to the optimal design.

6. DISCUSSION

In the evolving landscape of anti-cancer treatments, such as the emergence of high response rate treatments like cell and immuno-oncology therapies, the need for adaptive trial designs that can dynamically explore dose-response relationships is expanding. Therefore, in this paper, our goal is to propose a two-dose two-stage design, which extends on the popular traditional two-stage design proposed by Simon in 1989, that is more suitable for the modern advancements of anti-cancer treatments and more flexible in exploring the dose-response relationship in early-phase oncology treatment development.

In the proposed method, the flexibility of implementing the decision rules enables the design to accommodate the risk-benefit profile evaluation for two doses. Such adaptability is a key strength of our proposed method, providing a framework that aligns with the comprehensive nature of early-phase oncology therapy development.

In the proposed design, we acknowledge the potential high response rates of modern oncology therapy by incorporating early termination for efficacy in the first stage. This approach enables further sample size reduction and is aligned with the treatment capabilities of recent medicine compared to Simon's original design, which only included early termination for futility. Moreover, we introduce decision rules in the proposed method which enables our design to be flexible under different therapeutic outcomes when identifying the superior dose that optimizes the balance between therapeutic benefits and safety risks. By controlling type I error rigorously in the algorithm, our design also provides high confidence for the anti-tumor activities of the superior dose that the design identifies for the next phases of development. The numerical examples also illustrate the flexibility of our design in its application under varying scenarios.

While our method primarily centers around a single efficacy endpoint, ORR, we acknowledge that the evaluation of anti-tumor activities often encompasses a broader spectrum of metrics. Evidence for anti-tumor activities may involve multiple metrics such as the overall and partial response, response durability, overall survival, and relevant biomarkers. However, at the design stage, it is important to focus on a single parameter, which is different from the totality of evidence for treatment evaluation.

We recognize the limitation that the proposed method only selects one dose from the first stage among the two doses evaluated. In situations where there are no clear differences between the two doses in risk-benefit profiles in all available data, the lower dose is recommended moving forward to the next stage. In such situations, there might be reasons to expect a higher efficacy from the higher doses so that both doses will be moved to the next stage. Further research will be required for such study designs so that the

optimal sample size can be derived by controlling the type I error and achieving adequate power.

Moreover, although it is not common to establish quantitative targeting criteria for safety performance except for certain well-characterized safety endpoints such as dose-limiting toxicities at the design stage, the safety profiles such as serious adverse events, the pharmacokinetic parameters, and a comprehensive evaluation of the frequency, severity, duration, and the manageability of adverse events are also important factors to consider for dose selection. The benefit and risk ratio should also be assessed to ensure that the gain in efficacy for an increased dose will result in no substantial loss in safety.

APPENDIX A. TYPE I ERROR DERIVATION

A.1 Overall Type I Error

$$\begin{aligned}
& R_1 + R_{21} + R_{22} | \theta_1 \leq \theta_0 \text{ and } \theta_2 \leq \theta_0 \\
& = (1 - \text{Bin}(r_1 - 1 | \theta_1, n_1))(1 - \text{Bin}(r_1 - 1 | \theta_2, n_1)) \\
& + \text{Bin}(r_1 - 1 | \theta_1, n_1)(1 - \text{Bin}(r_1 - 1 | \theta_2, n_1)) \\
& + \text{Bin}(r_1 - 1 | \theta_2, n_1)(1 - \text{Bin}(r_1 - 1 | \theta_1, n_1)) \\
& + \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \text{Bin}(S_{11} | \theta_2, n_1) \\
& [1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
& + \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \text{Bin}(S_{12} - 1 | \theta_1, n_1) \\
& [1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)] \\
& = 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
& + \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \text{Bin}(S_{11} | \theta_2, n_1) \\
& [1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
& + \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \text{Bin}(S_{12} - 1 | \theta_1, n_1) \\
& [1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)] \\
& = 1 - \left(\sum_{k=0}^{r_1-1} \binom{n_1}{k} \theta_1^k (1 - \theta_1)^{n_1-k} \right) \\
& \left(\sum_{k=0}^{r_1-1} \binom{n_1}{k} \theta_2^k (1 - \theta_2)^{n_1-k} \right) \\
& + \sum_{S_{11}=a_1+1}^{r_1-1} \binom{n_1}{S_{11}} \theta_1^{S_{11}} (1 - \theta_1)^{n_1-S_{11}} \\
& \left(\sum_{k=0}^{S_{11}} \binom{n_1}{k} \theta_2^k (1 - \theta_2)^{n_1-k} \right) \\
& \left[1 - \sum_{k=0}^{r-S_{11}-1} \binom{n_2}{k} \theta_1^k (1 - \theta_1)^{n_2-k} \right]
\end{aligned}$$

$$\begin{aligned}
& + \sum_{S_{11}=a_1+1}^{r_1-1} \binom{n_1}{S_{11}} \theta_1^{S_{11}} (1 - \theta_1)^{n_1-S_{11}} \\
& \left(\sum_{k=0}^{S_{11}} \binom{n_1}{k} \theta_2^k (1 - \theta_2)^{n_1-k} \right) \\
& \left[1 - \sum_{k=0}^{r-S_{11}-1} \binom{n_2}{k} \theta_1^k (1 - \theta_1)^{n_2-k} \right].
\end{aligned}$$

A.2 Individual Type I Error

A.2.1 Maximum Type I Error for Dose I

$$\begin{aligned}
& P_{H_{01}}(S_{11} \geq r_1) \\
& + P_{H_{01}}(a_1 < S_{11} < r_1 \cap S_{11} \geq S_{12} \cap S \geq r) | \\
& \theta_1 \leq \theta_0 \text{ and } \theta_2 = 0 \\
& = 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \\
& + \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \text{Bin}(S_{11} | \theta_2, n_1) \\
& [1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
& \theta_1 \leq \theta_0 \text{ and } \theta_2 = 0 \\
& = 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \\
& + \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \\
& [1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
& \leq \alpha_0.
\end{aligned}$$

A.2.2 Maximum Type I Error for Dose II

$$\begin{aligned}
& P_{H_{02}}(S_{12} \geq r_1) \\
& + P_{H_{02}}(a_1 < S_{12} < r_1 \cap S_{12} > S_{11} \cap S \geq r) | \\
& \theta_1 = 0 \text{ and } \theta_2 \leq \theta_0 \\
& = 1 - \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
& + \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \text{Bin}(S_{12} - 1 | \theta_1, n_1) \\
& [1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)] \\
& \theta_1 = 0 \text{ and } \theta_2 \leq \theta_0 \\
& = 1 - \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
& + \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \\
& [1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)] \\
& \leq \alpha_0.
\end{aligned}$$

APPENDIX B. POWER DERIVATION

B.1 Overall Power

$$1 - \beta$$

$$\begin{aligned}
&= R_1 + R_{21} + R_{22} | \theta_1 = \theta_A \text{ or } \theta_2 = \theta_A \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
&+ \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \text{Bin}(S_{11} | \theta_2, n_1) \\
&[1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
&+ \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \text{Bin}(S_{12} - 1 | \theta_1, n_1) \\
&[1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)].
\end{aligned}$$

B.2 Power for Dose I at $(\theta_1 = \theta_A, \theta_2 = \theta_0)$

$$\begin{aligned}
&1 - \beta_1 \\
&= P_{H_{01}}(S_{11} \geq r_1) \\
&+ P_{H_{01}}(a_1 < S_{11} < r_1 \cap S_{11} \geq S_{12} \cap S \geq r) | \\
&\theta_1 = \theta_A, \theta_2 = \theta_0 \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \\
&+ \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_1, n_1) \text{Bin}(S_{11} | \theta_2, n_1) \\
&[1 - \text{Bin}(r - S_{11} - 1 | \theta_1, n_2)] \\
&\theta_1 = \theta_A, \theta_2 = \theta_0 \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_A, n_1) \\
&+ \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_A, n_1) \text{Bin}(S_{11} | \theta_0, n_1) \\
&[1 - \text{Bin}(r - S_{11} - 1 | \theta_A, n_2)].
\end{aligned}$$

B.3 Power for Dose II at $(\theta_1 = \theta_0, \theta_2 = \theta_A)$

$$\begin{aligned}
&1 - \beta_2 \\
&= P_{H_{02}}(S_{12} \geq r_1) \\
&+ P_{H_{02}}(a_1 < S_{12} < r_1 \cap S_{12} > S_{11} \cap S \geq r) | \\
&\theta_1 = \theta_0, \theta_2 = \theta_A \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
&+ \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_2, n_1) \text{Bin}(S_{12} | \theta_1, n_1) \\
&[1 - \text{Bin}(r - S_{12} - 1 | \theta_2, n_2)] \\
&\theta_1 = \theta_0, \theta_2 = \theta_A \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_A, n_1) \\
&+ \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_A, n_1) \text{Bin}(S_{12} - 1 | \theta_0, n_1) \\
&[1 - \text{Bin}(r - S_{12} - 1 | \theta_A, n_2)].
\end{aligned}$$

B.4 Power at $(\theta_1 = \theta_A, \theta_2 = \theta_A)$

$$1 - \beta_0$$

$$\begin{aligned}
&= R_1 + R_{21} + R_{22} | \theta_1 = \theta_A, \theta_2 = \theta_A \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_A, n_1) \text{Bin}(r_1 - 1 | \theta_A, n_1) \\
&+ \sum_{S_{11}=a_1+1}^{r_1-1} \text{bin}(S_{11} | \theta_A, n_1) \text{Bin}(S_{11} | \theta_A, n_1) \\
&[1 - \text{Bin}(r - S_{11} - 1 | \theta_A, n_2)] \\
&+ \sum_{S_{12}=a_1+1}^{r_1-1} \text{bin}(S_{12} | \theta_A, n_1) \text{Bin}(S_{12} - 1 | \theta_A, n_1) \\
&[1 - \text{Bin}(r - S_{12} - 1 | \theta_A, n_2)].
\end{aligned}$$

APPENDIX C. PET DERIVATION

C.1 PET at Null Parameter Space

$$\begin{aligned}
&R_1 + A_1 | \theta_1 = \theta_0 \text{ and } \theta_2 = \theta_0 \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
&+ \text{Bin}(a_1 | \theta_1, n_1) \text{Bin}(a_1 | \theta_2, n_1) | \\
&\theta_1 = \theta_0 \text{ and } \theta_2 = \theta_0 \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_0, n_1)^2 + \text{Bin}(a_1 | \theta_0, n_1)^2.
\end{aligned}$$

C.2 PET at Alternative Parameter Space

$$\begin{aligned}
&R_1 + A_1 | \theta_1 = \theta_A \text{ and } \theta_2 = \theta_A \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_1, n_1) \text{Bin}(r_1 - 1 | \theta_2, n_1) \\
&+ \text{Bin}(a_1 | \theta_1, n_1) \text{Bin}(a_1 | \theta_2, n_1) | \\
&\theta_1 = \theta_A \text{ and } \theta_2 = \theta_A \\
&= 1 - \text{Bin}(r_1 - 1 | \theta_A, n_1)^2 + \text{Bin}(a_1 | \theta_A, n_1)^2.
\end{aligned}$$

APPENDIX D. RESULTS

D.1 Complete Results of the Optimal and Minimax Design when Powering at $\theta_1 = \theta_A$ and $\theta_2 = \theta_A$

Table D.1 shows the results of designs that satisfy the type I error and required power for various design parameters θ_0 and θ_A , when powering at $\theta_1 = \theta_A$ and $\theta_2 = \theta_A$, the same targeting anti-tumor activities for both doses, for relative treatment differences $\theta_A - \theta_0$ of either 0.3 or 0.2. Table 1 includes the total sample size, n , the sample size of the first stage n_1 , the sample size of the second stage n_2 , the lower a_1 and upper r_1 thresholds of responses for terminating the study at Stage I, the upper threshold of the total responses for claiming efficacy from both stages r , the Type I error, the power, as well as the probability of early terminations PET_1 at null parameter space, PET_2 at alternative parameter space and PET_{Avg} which is the average of PET_1 and PET_2 , and the corresponding expected sample size EN_1 , EN_2 and EN_{Avg} . The study designs presented in Table D.1 power the alternative points at the minimum of 80% and control the type I error at the level of 1-sided 0.05.

Table D.1. Power the Region of $\theta_1 \geq \theta_A$ and $\theta_2 \geq \theta_A$.

Difference	Parameter	n	n_1	n_2	a_1	r_1	r	Type I Err	Power	PET ₁	PET ₂	PET _{Avg}	EN ₁	EN ₂	EN _{Avg}
0.3	$\theta_0 = 0.2, \theta_A = 0.5$	20	6	8	1	4	7	0.05	0.81	0.46	0.58	0.52	17	16	16
	$\theta_0 = 0.3, \theta_A = 0.6$	23	6	11	1	5	10	0.05	0.80	0.20	0.41	0.31	21	19	20
		21	7	7	3	6	8	0.05	0.81	0.77	0.38	0.57	16	19	17
	$\theta_0 = 0.4, \theta_A = 0.7$	24	7	10	3	6	12	0.05	0.80	0.54	0.57	0.55	19	19	19
	$\theta_0 = 0.5, \theta_A = 0.8$	22	7	8	4	7	12	0.04	0.81	0.61	0.40	0.51	18	19	18
0.2	$\theta_0 = 0.2, \theta_A = 0.4$	41	11	19	3	6	11	0.05	0.81	0.73	0.52	0.62	28	32	30
		54	14	26	5	9	19	0.04	0.81	0.63	0.42	0.52	29	30	29
	$\theta_0 = 0.3, \theta_A = 0.5$	53	15	23	6	9	18	0.05	0.81	0.79	0.61	0.70	35	40	37
		52	20	12	8	11	16	0.05	0.80	0.82	0.72	0.77	43	44	43
		59	15	29	8	11	24	0.05	0.80	0.84	0.54	0.69	35	44	40
	$\theta_0 = 0.4, \theta_A = 0.6$	58	16	26	8	12	24	0.04	0.80	0.75	0.39	0.57	39	48	44
		56	17	22	9	12	22	0.05	0.80	0.85	0.59	0.72	38	44	41
		55	21	13	11	14	20	0.04	0.80	0.86	0.67	0.77	44	47	46
		56	15	26	9	12	28	0.05	0.80	0.76	0.58	0.67	37	41	39
	$\theta_0 = 0.5, \theta_A = 0.7$	55	16	23	10	13	26	0.05	0.81	0.82	0.55	0.68	37	43	40
		53	19	15	12	15	23	0.05	0.81	0.84	0.11	0.48	41	52	46

Table D.2. Power the Region of $\theta_1 \geq \theta_A$ or $\theta_2 \geq \theta_A$.

Difference	Parameter	n	n_1	n_2	a_1	r_1	r	Type I Err	Power ₀	Power ₁	Power ₂	PET ₁	PET ₂	PET _{Avg}	EN ₁	EN ₂	EN _{Avg}
0.3	$\theta_0 = 0.2, \theta_A = 0.5$	37	10	17	1	6	11	0.05	0.95	0.85	0.81	0.15	0.61	0.38	35	27	31
		37	10	17	2	6	11	0.04	0.95	0.84	0.80	0.47	0.61	0.54	29	27	28
		35	11	13	2	6	10	0.05	0.95	0.84	0.81	0.40	0.75	0.58	30	26	28
	$\theta_0 = 0.3, \theta_A = 0.6$	41	10	21	2	7	16	0.05	0.95	0.85	0.80	0.17	0.62	0.39	38	29	33
		39	12	15	4	8	14	0.05	0.95	0.83	0.80	0.54	0.69	0.62	31	29	30
	$\theta_0 = 0.4, \theta_A = 0.7$	42	11	20	5	9	19	0.04	0.96	0.84	0.80	0.58	0.53	0.56	31	32	31
	$\theta_0 = 0.5, \theta_A = 0.8$	39	10	19	5	9	21	0.05	0.96	0.86	0.82	0.41	0.61	0.51	32	28	30
37	11	15	6	10	19	0.05	0.95	0.84	0.81	0.54	0.54	0.54	29	29	29		
0.2	$\theta_0 = 0.2, \theta_A = 0.4$	79	22	35	5	10	19	0.05	0.95	0.83	0.81	0.55	0.62	0.58	60	58	59
		77	25	27	6	10	18	0.05	0.95	0.81	0.80	0.64	0.83	0.73	60	55	58
		94	24	46	7	13	31	0.05	0.95	0.83	0.81	0.34	0.66	0.50	79	64	71
	$\theta_0 = 0.3, \theta_A = 0.5$	93	25	43	8	14	30	0.05	0.95	0.82	0.80	0.47	0.57	0.52	73	69	71
		92	27	38	9	14	29	0.05	0.95	0.82	0.80	0.56	0.75	0.66	71	64	68
		99	26	47	10	17	40	0.05	0.94	0.82	0.80	0.29	0.60	0.44	86	71	79
	$\theta_0 = 0.4, \theta_A = 0.6$	98	32	34	15	20	36	0.05	0.95	0.81	0.80	0.71	0.72	0.72	74	74	74
		97	38	21	18	22	33	0.05	0.95	0.80	0.80	0.78	0.90	0.84	81	79	80
		96	25	46	13	19	46	0.05	0.95	0.83	0.80	0.44	0.57	0.51	76	70	73
		95	27	41	15	20	44	0.05	0.95	0.82	0.80	0.63	0.66	0.64	70	68	69
	$\theta_0 = 0.5, \theta_A = 0.7$	94	33	28	19	23	40	0.05	0.95	0.81	0.80	0.76	0.85	0.80	73	71	72

D.2 Complete Results of the Optimal and Minimax Design when Powering at $\theta_1 = \theta_A$ or $\theta_2 = \theta_A$

Table D.2 shows the results of design when powering at $\theta_1 = \theta_A$ or $\theta_2 = \theta_A$, the same targeting anti-tumor activities for both doses, for relative treatment differences $\theta_A - \theta_0$ of either 0.3 or 0.2. In such designs, the studies are powered when either dose or both doses have response rates θ_A . Table D.2 includes the total sample size, n , the sample size of the first stage n_1 , the sample size of the second stage n_2 , the lower a_1 and upper r_1 thresholds of responses for termi-

nating the study at Stage I, the upper threshold of the total responses for claiming efficacy from both stages r , the Type I error, power at $(\theta_1 = \theta_A, \theta_2 = \theta_A)$, power for dose I at $(\theta_1 = \theta_A, \theta_2 = \theta_0)$, power for dose II at $(\theta_1 = \theta_0, \theta_2 = \theta_A)$ as well as the probability of early terminations PET_1 at null parameter space, PET_2 at alternative parameter space and PET_{Avg} which is the average of PET_1 and PET_2 , and the corresponding expected sample size EN_1 , EN_2 and EN_{Avg} . The study designs presented in Table D.2 power the alternative points at the minimum of 80% and control the type I error at the level of 1-sided 0.05.

Accepted 24 January 2025

REFERENCES

- [1] BARTROFF, J., LAI, T. L. and NARASIMHAN, B. (2014). A new approach to designing phase I-II cancer trials for cytotoxic chemotherapies. *Statistics in medicine* **33**(16) 2718–2735. <https://doi.org/10.1002/sim.6124>. MR3256534
- [2] CHEKOL ABEBE, E., YIBELTAL SHIFERAW, M., TADELE ADMASU, F. and ASMAMAW DEJENIE, T. (2022). Ciltacabtagene autoleucel: The second anti-BCMA CAR T-cell therapeutic armamentarium of relapsed or refractory multiple myeloma. *Frontiers in immunology* **13** 991092.
- [3] CHEN, T. T. (1997). Optimal three-stage designs for phase II cancer clinical trials. *Statistics in medicine* **16**(23) 2701–2711.
- [4] CRIPPA, A., DISCACCIATI, A., BOTTAI, M., SPIEGELMAN, D. and ORSINI, N. (2019). One-stage dose–response meta-analysis for aggregated data. *Statistical methods in medical research* **28**(5) 1579–1596. <https://doi.org/10.1177/0962280218773122>. MR3941095
- [5] JI, Y., LIU, P., LI, Y. and NEBIYOU BEKELE, B. (2010). A modified toxicity probability interval method for dose-finding trials. *Clinical trials* **7**(6) 653–663.
- [6] LEE, S. Y. (2023). A flexible dose-response modeling framework based on continuous toxicity outcomes in phase I cancer clinical trials. *Trials* **24**(1) 745.
- [7] LIN, Y. and SHIH, W. J. (2004). Adaptive two-stage designs for single-arm phase IIA cancer clinical trials. *Biometrics* **60**(2) 482–490. <https://doi.org/10.1111/j.0006-341X.2004.00193.x>. MR2067004
- [8] LIU, S. and YUAN, Y. (2015). Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal Statistical Society: Series C: Applied Statistics* 507–523. <https://doi.org/10.1111/rssc.12089>. MR3325461
- [9] PARASHAR, D., BOWDEN, J., STARR, C., WERNISCH, L. and MANDER, A. (2016). An optimal stratified Simon two-stage design. *Pharmaceutical statistics* **15**(4) 333–340.
- [10] SIMON, R. (1989). Optimal two-stage designs for phase II clinical trials. *Controlled clinical trials* **10**(1) 1–10.
- [11] US FOOD AND DRUG ADMINISTRATION (2022). *Expansion cohorts: use in first-in-human clinical trials to expedite development of oncology drugs and biologics*. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/expansion-cohorts-use-first-human-clinical-trials-expedite-development-oncology-drugs-and-biologics> Accessed 2023-12-01.
- [12] US FOOD AND DRUG ADMINISTRATION (2023). *Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases*. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/optimizing-dosage-human-prescription-drugs-and-biological-products-treatment-oncologic-diseases> Accessed 2023-12-01.
- [13] WHITEHEAD, J. (2014). One-stage and two-stage designs for phase II clinical trials with survival endpoints. *Statistics in medicine* **33**(22) 3830–3843. <https://doi.org/10.1002/sim.6196>. MR3260663
- [14] YE, F. and SHYR, Y. (2007). Balanced two-stage designs for phase II clinical trials. *Clinical Trials* **4**(5) 514–524.
- [15] ZOHAR, S. and CHEVRET, S. (2003). Phase I (or phase II) dose-ranging clinical trials: proposal of a two-stage Bayesian design. *Journal of Biopharmaceutical Statistics* **13**(1) 87–101. <https://doi.org/10.1002/sim.6338>. MR3286234

Meiruo Xiang. Storrs, CT, University of Connecticut, United States. E-mail address: meiruo.xiang@uconn.edu

Yangruijie Ma. Chicago, IL, Division of Biostatistics, Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, United States. E-mail address: anna.myrj@gmail.com

Qian Li. Bethesda, MD, StatsVita, LLC, United States. E-mail address: qianhelenli@gmail.com