Design and Sample Size Determination for Multiple-dose Randomized Phase II Trials for Dose Optimization

Peng Yang

Rice University

June 5, 2023

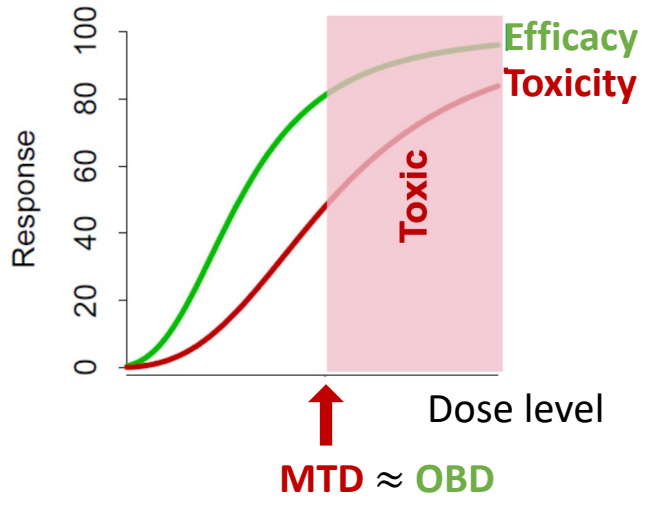
Joint work with Daniel Li (BMS), Ruitao Lin (MD Anderson), Bo Huang (Pfizer), and Ying Yuan (MD Anderson)

Design and Sample Size Determination for Multi-Dose Randomized Trial

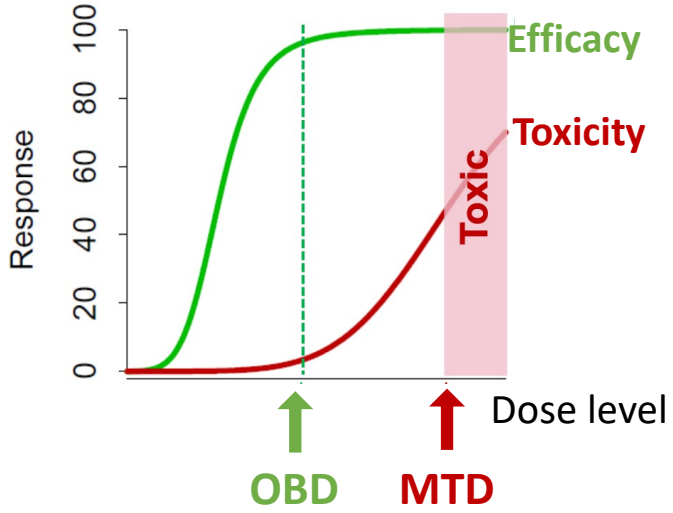
Background – Project Optimus

- In 2022, FDA initiated Project Optimus "to reform the dose optimization and dose selection paradigm in oncology drug development."
- Paradigm shifting from maximum tolerated dose (MTD) to optimal biological dose (OBD).

Conventional Chemotherapy



Targeted Therapies

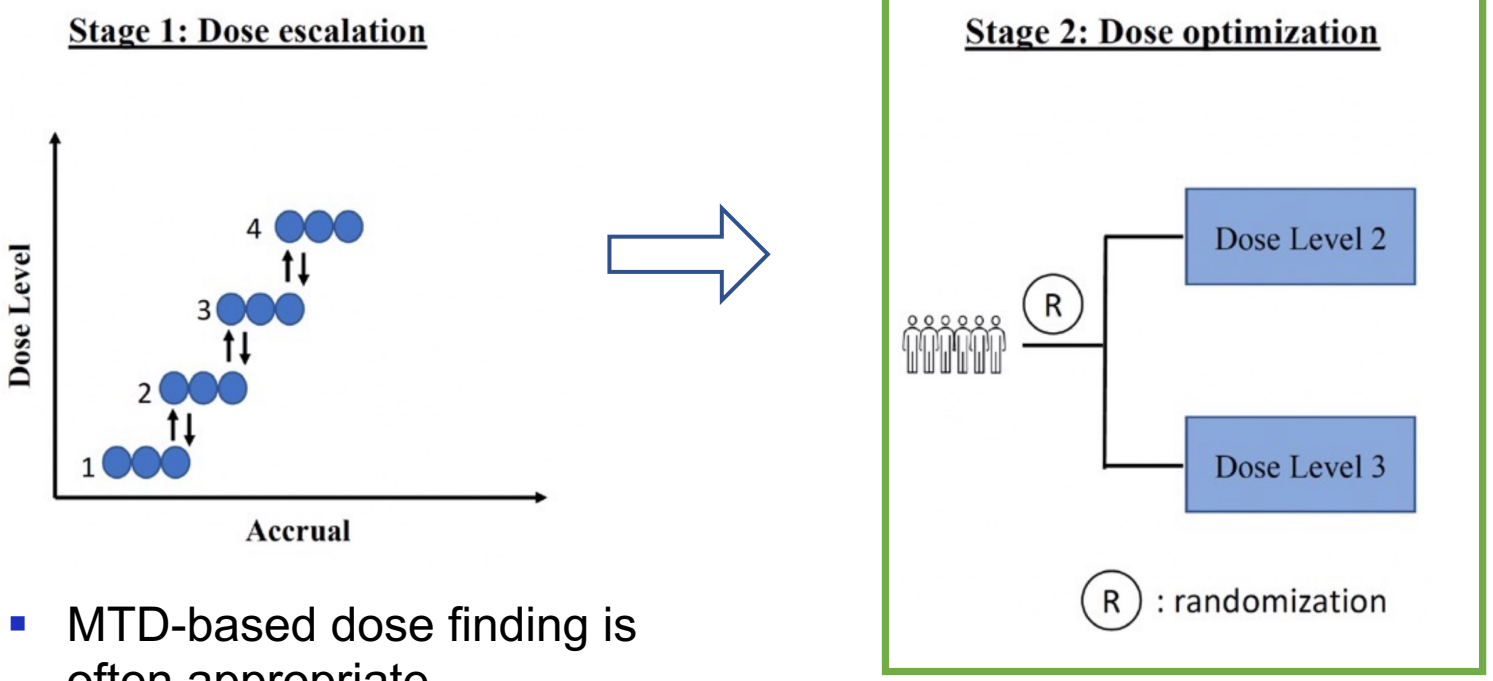


- MTD-based dose finding is often appropriate to inform RP2D
- Safety alone is not sufficient to inform optimal RP2D

Design and Sample Size Determination for Multi-Dose Randomized Trial

Background – Design strategies to find OBD

- Trial designs to compare multiple dosages (FDA, 2023)
  - The trial should be sized to allow for sufficient assessment of activity, safety, and tolerability for each dosage.
  - The trial does not need to be powered to demonstrate statistical superiority of a dosage or statistical non-inferiority among the dosages



- MTD-based dose finding is often appropriate

Design and Sample Size Determination for Multi-Dose Randomized Trial

Background – Two-stage decision-making paradigm

1

Determination of OBD admissible dose set

- The objective is to identify a dose set that satisfy certain safety and efficacy requirements (i.e., OBD admissible set ***A***).

2

Identification of the OBD

- The OBD will be selected from A based on the totality of activity, safety and tolerability data.
  - "Relevant nonclinical and clinical data, as well as the dose- and exposure- response relationships for safety and efficacy should be evaluated to select a dosage(s) for clinical trial(s)" (FDA guidance)
  - Unlikely/impossible to formulate statistical decision rules to capture all quantitative and qualitative considerations relevant to the final OBD selection

Design and Sample Size Determination for Multi-Dose Randomized Trial

Background – Two-stage decision-making paradigm

1

Determination of OBD admissible dose set

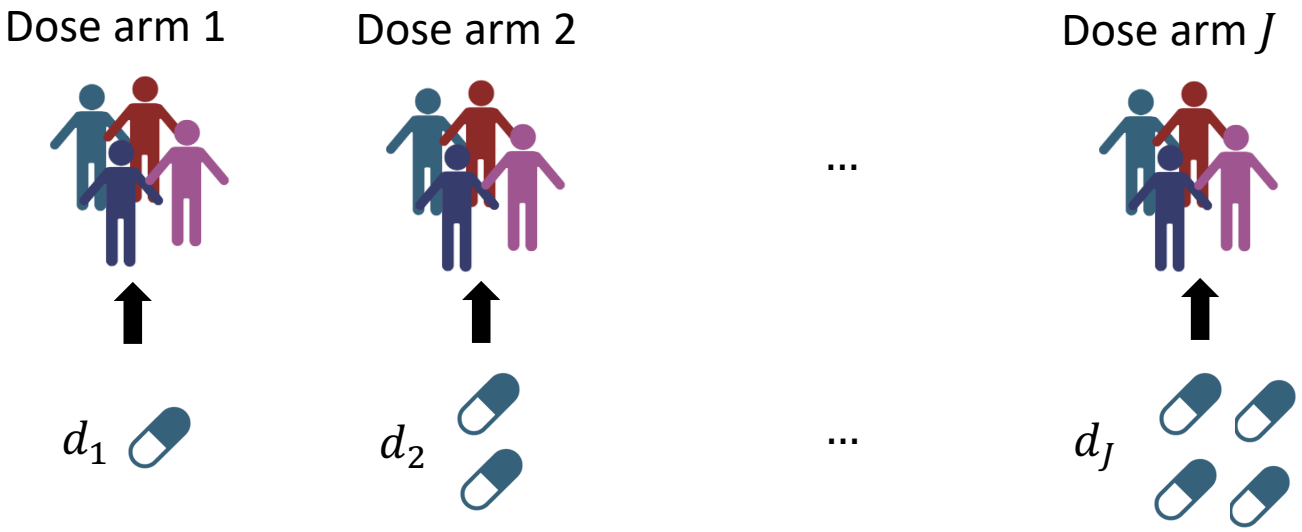
- The objective is to identify a dose set that satisfy certain safety and efficacy requirements (i.e., OBD admissible set ***A***).

Goal: formulize the deign in this step to ensure the trial satisfies certain statistical properties, including type I error and power.

Design and Sample Size Determination for Multi-Dose Randomized Trial

Statistical modelling – set ups

- Consider a multiple-dose randomized trial, where a total of $J \times n$ patients are equally randomized to J doses, $d_1 < d_2 < \cdots < d_J$.
  - In most applications, J = 2 or 3, and d_J is often the MTD or maximum administered dose.

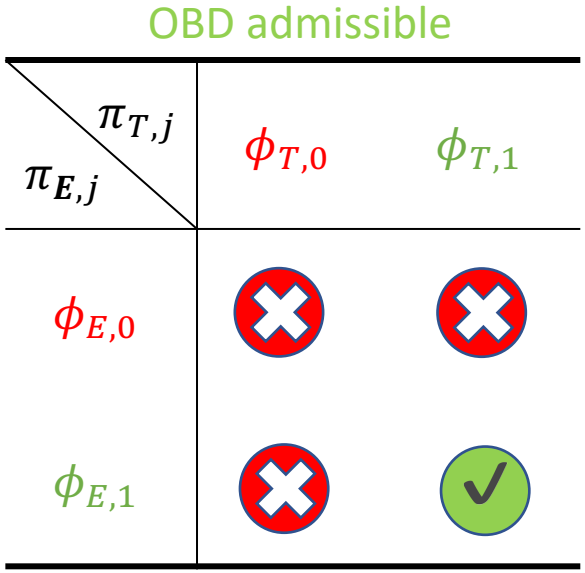


- Let Y_T and Y_E denote the binary toxicity and efficacy endpoints, respectively.
  - Example of Y_T: dose-limiting toxicity, dichotomized total toxicity burden.
  - Example of Y_E: objective responses, efficacy surrogate endpoints.

Design and Sample Size Determination for Multi-Dose Randomized Trial

Statistical modelling – assumptions

- Let $\pi_{T,j} = \Pr(Y_T = 1|d_j)$ and $\pi_{E,j} = \Pr(Y_E = 1|d_j)$ denotes the probability of the occurrence of toxicity and efficacy events.
- Assuming that $\pi_{T,j}$ and $\pi_{E,j}$ are **non-decreasing** with respect to the increasing of dose levels.
  - Randomized dose optimization trials have the same drug but with **ordered doses**
- For toxicity endpoint, we assume
  - $\phi_{T,0}$: the **null toxicity** rate that is **high** and deemed unacceptable;
  - $\phi_{T,1}$: the **alternative toxicity rate** that is **low** and deemed acceptable;
- For efficacy endpoint, we let
  - $\phi_{E,0}$: the **null efficacy** rate that is **low** and deemed unacceptable;
  - $\phi_{E,1}$: the **alternative efficacy** rate that is **high** and deemed acceptable



Design and Sample Size Determination for Multi-Dose Randomized Trial

Statistical modelling – global type I error

- Consider H_0: None of the doses is the OBD.
  - **Challenge**: consists of multiple hypotheses ($\sum_{i=1}^{J+1} j$).

$$H_0(s,k): \underbrace{\pi_{T,1} = \pi_{T,2} = \cdots = \pi_{T,s} = \phi_{T,1}}_{\text{acceptable toxicity}} < \underbrace{\pi_{T,s+1} = \cdots = \pi_{T,k} = \pi_{T,k+1} = \cdots = \pi_{T,J} = \phi_{T,0}}_{\text{unacceptable toxicity}};$$

$$\underbrace{\pi_{E,1} = \pi_{E,2} = \cdots = \pi_{E,s} = \pi_{E,s+1} = \cdots = \pi_{E,k} = \phi_{E,0}}_{\text{unacceptable efficacy}} < \underbrace{\pi_{E,k+1} = \cdots = \pi_{E,J} = \phi_{E,1}}_{\text{acceptable efficacy}},$$

where $s, k \in \{0,1,\ldots,J\}$ with $s \leq k$.

		d_1	d_2			d_1	d_2
$H_0(0,0)$	Tox	$\phi_{T,0}$	$\phi_{T,0}$	$H_0(0,0)$	Tox	$\phi_{T,1}$	$\phi_{T,0}$
	Eff	$\phi_{E,1}$	$\phi_{E,1}$		Eff	$\phi_{E,0}$	$\phi_{E,1}$
$H_0(0,1)$	Tox	$\phi_{T,0}$	$\phi_{T,0}$	$H_0(0,1)$	Tox	$\phi_{T,1}$	$\phi_{T,0}$
	Eff	$\phi_{E,0}$	$\phi_{E,1}$		Eff	$\phi_{E,0}$	$\phi_{E,0}$
$H_0(0,2)$	Tox	$\phi_{T,0}$	$\phi_{T,0}$	$H_0(0,2)$	Tox	$\phi_{T,1}$	$\phi_{T,1}$
	Eff	$\phi_{E,0}$	$\phi_{E,0}$		Eff	$\phi_{E,0}$	$\phi_{E,0}$

Design and Sample Size Determination for Multi-Dose Randomized Trial

Statistical modelling – global type I error

- Consider H_0: None of the doses is the OBD.
  - **Challenge**: consists of multiple hypotheses ($\sum_{j=1}^{J+1} j$).

$$H_0(s,k): \underbrace{\pi_{T,1} = \pi_{T,2} = \cdots = \pi_{T,s} = \phi_{T,1}}_{\text{acceptable toxicity}} < \underbrace{\pi_{T,s+1} = \cdots = \pi_{T,k} = \pi_{T,k+1} = \cdots = \pi_{T,J} = \phi_{T,0}}_{\text{unacceptable toxicity}};$$

$$\underbrace{\pi_{E,1} = \pi_{E,2} = \cdots = \pi_{E,s} = \pi_{E,s+1} = \cdots = \pi_{E,k} = \phi_{E,0}}_{\text{unacceptable efficacy}} < \underbrace{\pi_{E,k+1} = \cdots = \pi_{E,J} = \phi_{E,1}}_{\text{acceptable efficacy}},$$

where $s, k \in \{0,1,\ldots,J\}$ with $s \leq k$.

- Define *global type I error* that encompasses all $H_0(s,k)$ as follows:

$$\alpha = \Pr(\text{reject } H_0 | H_0)$$

$$= \max_{0 \leq s \leq J,\ s \leq k \leq J} \{\alpha(s,k)\}.$$

where $\alpha(s,k) = \Pr(\text{reject } H_0(s,k) | H_0(s,k))$.

Design and Sample Size Determination for Multi-Dose Randomized Trial

Statistical modelling – generalized powers

- Consider H_1: At lease one does is the OBD admissible.
  - **Challenge**: consists of multiple hypotheses ($\sum_{j=1}^{J} j$).

$$H_1(u,v): \underbrace{\pi_{T,1} = \pi_{T,2} = \cdots = \pi_{T,u} = \pi_{T,u+1} = \cdots = \pi_{T,v} = \phi_{T,1}}_{\text{acceptable toxicity}} < \underbrace{\pi_{T,v+1} = \cdots = \pi_{T,J} = \phi_{T,0}}_{\text{unacceptable toxicity}};$$

$$\underbrace{\pi_{E,1} = \pi_{E,2} = \cdots = \pi_{E,u} = \phi_{E,0}}_{\text{unacceptable efficacy}} < \underbrace{\pi_{E,u+1} = \cdots = \pi_{E,v} = \pi_{E,v+1} = \cdots = \pi_{E,J} = \phi_{E,1}}_{\text{acceptable efficacy}},$$

where $u, v \in \{0,1,\ldots,J\}$ with $u < v$.

		d_1	d_2
$H_1(0,1)$	Tox	$\phi_{T,1}$	$\phi_{T,0}$
	Eff	$\phi_{E,1}$	$\phi_{E,1}$
$H_1(0,2)$	Tox	$\phi_{T,1}$	$\phi_{T,1}$
	Eff	$\phi_{E,1}$	$\phi_{E,1}$
$H_1(1,2)$	Tox	$\phi_{T,1}$	$\phi_{T,1}$
	Eff	$\phi_{E,0}$	$\phi_{E,1}$

: OBD admissible

Statistical modelling – generalized powers

- Consider H_1: At lease one does is the OBD admissible.
  - **Challenge**: consists of multiple hypotheses ($\sum_{j=1}^{J} j$).
  - **Additional challenge**: the standard definition of power, which reject the H_0, is not sufficient to account for the characteristics of dose optimization.

		d_1	d_2
$H_1(0,1)$	Tox	$\phi_{T,1}$	$\phi_{T,0}$
	Eff	$\phi_{E,1}$	$\phi_{E,1}$
Selection		✓	
			✓
		✓	✓

- It is important to account for the quality of the admissible dose selection!

Design and Sample Size Determination for Multi-Dose Randomized Trial

Statistical modelling – generalized powers

- *Generalized power I*:

$$\beta_1(u,v) = \Pr\left(\begin{array}{c}\text{reject } H_0 \text{ and all dose in } A \text{ are truly} \\ \text{safe and efficacious}\end{array} \,\middle|\, H_1(u,v)\right).$$

where *A* denotes the admissible dose set selected by the design.

- *Generalized power II*:

$$\beta_2(u,v) = \Pr\left(\begin{array}{c}\text{reject } H_0 \text{ and at leaset one dose in } A \text{ are truly} \\ \text{safe and efficacious}\end{array} \,\middle|\, H_1(u,v)\right).$$

	β_1	β_2
Tolerate false positive	✗	✓

- Both generalized powers are stricter than the standard power.
- The choice of power depends on the characteristics of the trial and user's tolerability of false positive.
- Power II is a good option when reducing the sample size is of top priority.

Statistical modelling – generalized powers

- Consider H_1: At lease one does is the OBD admissible.
  - **Challenge**: consists of multiple hypotheses ($\sum_{j=1}^{J} j$).

$$H_1(u,v): \underbrace{\pi_{T,1} = \pi_{T,2} = \cdots = \pi_{T,u} = \pi_{T,u+1} = \cdots = \pi_{T,v} = \phi_{T,1}}_{\text{acceptable toxicity}} < \underbrace{\pi_{T,v+1} = \cdots = \pi_{T,J} = \phi_{T,0}}_{\text{unacceptable toxicity}};$$

$$\underbrace{\pi_{E,1} = \pi_{E,2} = \cdots = \pi_{E,u} = \phi_{E,0}}_{\text{unacceptable efficacy}} < \underbrace{\pi_{E,u+1} = \cdots = \pi_{E,v} = \pi_{E,v+1} = \cdots = \pi_{E,J} = \phi_{E,1}}_{\text{acceptable efficacy}},$$

where $u, v \in \{0,1,\ldots,J\}$ with $u < v$.

- We define *global power I and power II* that encompasses all $H_1(u,v)$ as follows

$$\beta_i = \min_{u,v \in \{0,\cdots,J\}, u<v} \beta_i(u,v), \text{ for } i = 1,2.$$

Design and Sample Size Determination for Multi-Dose Randomized Trial

Statistical modelling – least favorable set

Theorem 1. *Define the least favorable set* $\widetilde{H}_1 = \{H_1(j), j = 1, \cdots, J\}$, *where*

$$H_1(j) = \left(\begin{array}{ccc} \underbrace{\pi_{T,1} = \cdots = \pi_{T,j-1} = \phi_{T,1}}_{\textit{safe but futile}} & \underbrace{\pi_{T,j} = \phi_{T,1}}_{\textit{safe and efficacious}} & \underbrace{\pi_{T,j+1} = \cdots = \pi_{T,J} = \phi_{T,0}}_{\textit{toxic and efficacious}} \\ \underbrace{\pi_{E,1} = \cdots = \pi_{E,j-1} = \phi_{E,0}}_{\textit{safe but futile}} & \underbrace{\pi_{E,j} = \phi_{E,1}}_{\textit{safe and efficacious}} & \underbrace{\pi_{E,j+1} = \cdots = \pi_{E,J} = \phi_{E,1}}_{\textit{toxic and efficacious}} \end{array} \right).$$

For any $H_1(u, v)$, *with* $u, v \in \{0, 1, 2, \ldots, J\}$ *and* $u < v$, *there exists an* $H_1(j)$ *such that* $\beta_i(j) \leq \beta_i(u, v)$, $i = 1, 2$, *where* $\beta_1(j)$ *and* $\beta_2(j)$ *denote the generalized power I and II under* $H_1(j)$, *respectively.*

- The global power can be equivalently defined as follows

$$\beta_i = \min_{j \in \{1, \ldots, J\}} \beta_i(j) \text{ for } i = 1, 2.$$

Design and Sample Size Determination for Multi-Dose Randomized Trial

MERIT Design

- MERIT (Multiple-dosE RandomIzed Phase-II Trial)
  - Specify target global type I error α^* and power β^*;
  - Randomize $J \times n$ patients equally to J dosing arm;
  - In any dose arm d_j, if $n_{E,j} \geq m_E$ and $n_{T,j} \leq m_T$, we reject H_0 and claim that d_j is OBD admissible, where m_T and m_E are decision boundaries for toxicity and efficacy, respectively.

*$n_{E,j}$ and $n_{T,j}$ are the total number of patients who experience efficacy and toxicity events in dose arm d_j.

Optimal Design Parameters

- The optimal design parameters (n, m_T, m_E) are determined through numerical search such that the design controls the global type I error and power at nominal values α^* and β^*.
- Type I error

$$\begin{aligned}\alpha(s,k) &= \Pr(\text{reject } H_0(s,k) | H_0(s,k)) \\ &= 1 - \Big\{ \big(1 - \Pr(n_T \le m_T, n_E \ge m_E; n, \phi_{T,1}, \phi_{E,0})\big)^s \\ &\quad \times \big(1 - \Pr(n_T \le m_T, n_E \ge m_E; n, \phi_{T,0}, \phi_{E,0})\big)^{k-s} \\ &\quad \times \big(1 - \Pr(n_T \le m_T, n_E \ge m_E; n, \phi_{T,0}, \phi_{E,1})\big)^{J-k} \Big\}\end{aligned}$$

- Power

$$\begin{aligned}\beta_1(j) = \Pr(&n_{E,1} < m_E, \ldots, n_{E,j-1} < m_E, n_{T,j+1} > m_T, \ldots, n_{T,j} > m_T, \\ &n_{E,j} \ge m_E, n_{T,j} \le m_T \,|H_1(j)),\end{aligned}$$

$$\beta_2(j) = \Pr(n_{E,j} \ge m_E, n_{T,j} \le m_T \,|H_1(j))$$

Design and Sample Size Determination for Multi-Dose Randomized Trial

Practice Consideration

- In small samples, isotonic transformed $\{\tilde{n}_{T,j}, \tilde{n}_{E,j}\}$ should be applied before comparing with boundaries (m_T, m_E) when non-decreasing assumption is sound for toxicity and efficacy.
- In some trials, it maybe desirable to add futility and safety interim monitoring
  - Stop arm j for safety if $\Pr(\pi_{T,j} > \phi_{T,1} | \text{data}) > C_T$,
  - Stop arm j for futility if $\Pr(\pi_{E,j} < \phi_{E,1} | \text{data}) > C_E$,

  where C_T and C_E are probability cutoffs (e.g. 0.95).
- Whether to including interim monitoring depends on the availability of Y_T and Y_E, and other considerations. Typically, one or two interims are sufficient.

Design and Sample Size Determination for Multi-Dose Randomized Trial

Optimal Design Parameters

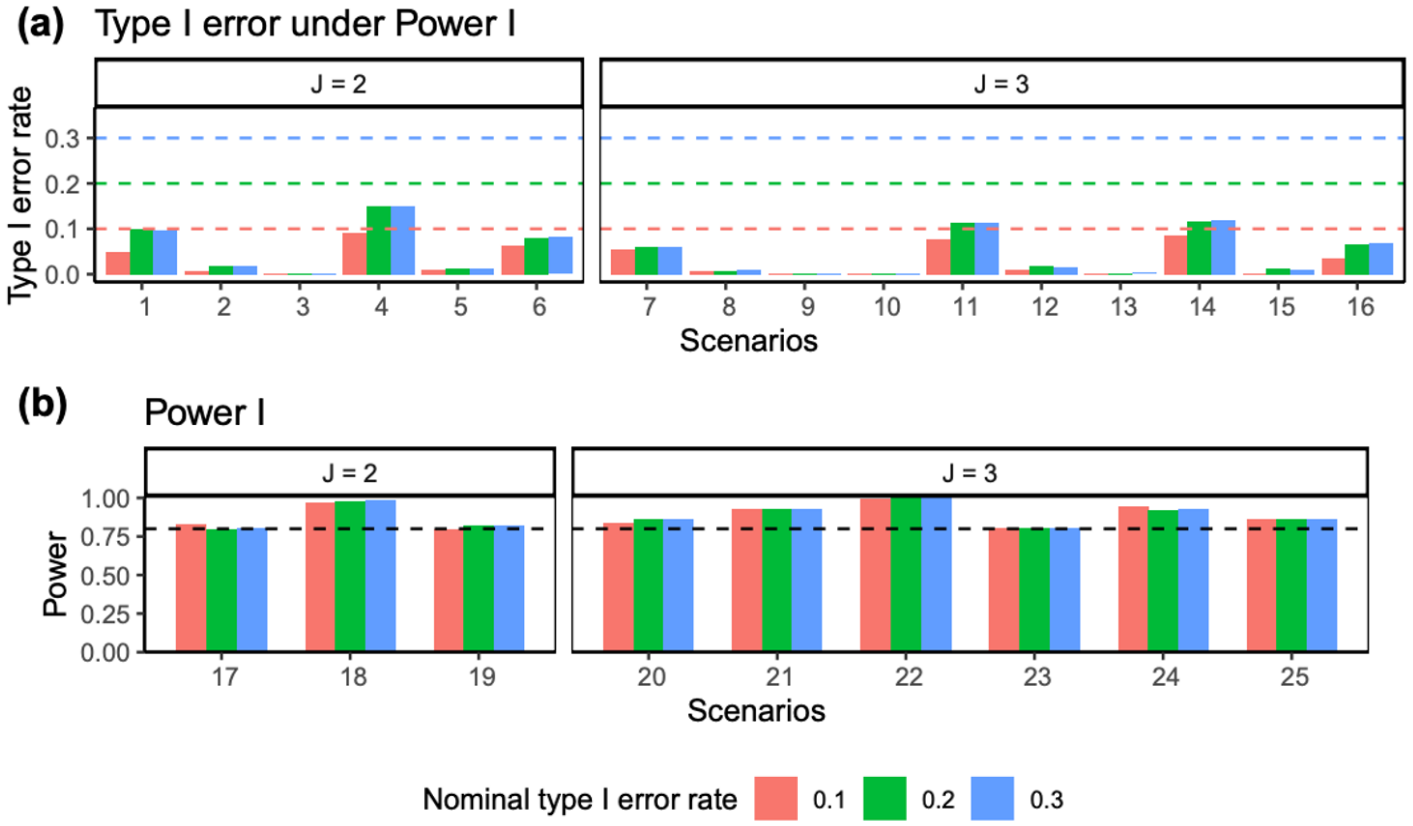
- Optimal design parameters (n, m_T, m_E) of MERIT design, when $J = 2$, $(\phi_{T,0}, \phi_{T,1}) = (0.4, 0.2)$, and $(\phi_{E,0}, \phi_{E,1}) = (0.2, 0.4)$.

β^*	β_1						β_2					
	$\alpha^* = 0.1$			$\alpha^* = 0.2$			$\alpha^* = 0.1$			$\alpha^* = 0.2$		
	n	m_T	m_E	n	m_T	m_E	n	m_T	m_E	n	m_T	m_E
0.6	26	7	6	23	6	26	25	6	5	18	5	4
0.7	33	9	7	30	8	33	33	8	6	24	7	5
0.8	44	12	8	39	11	44	39	11	8	30	8	5

Design and Sample Size Determination for Multi-Dose Randomized Trial

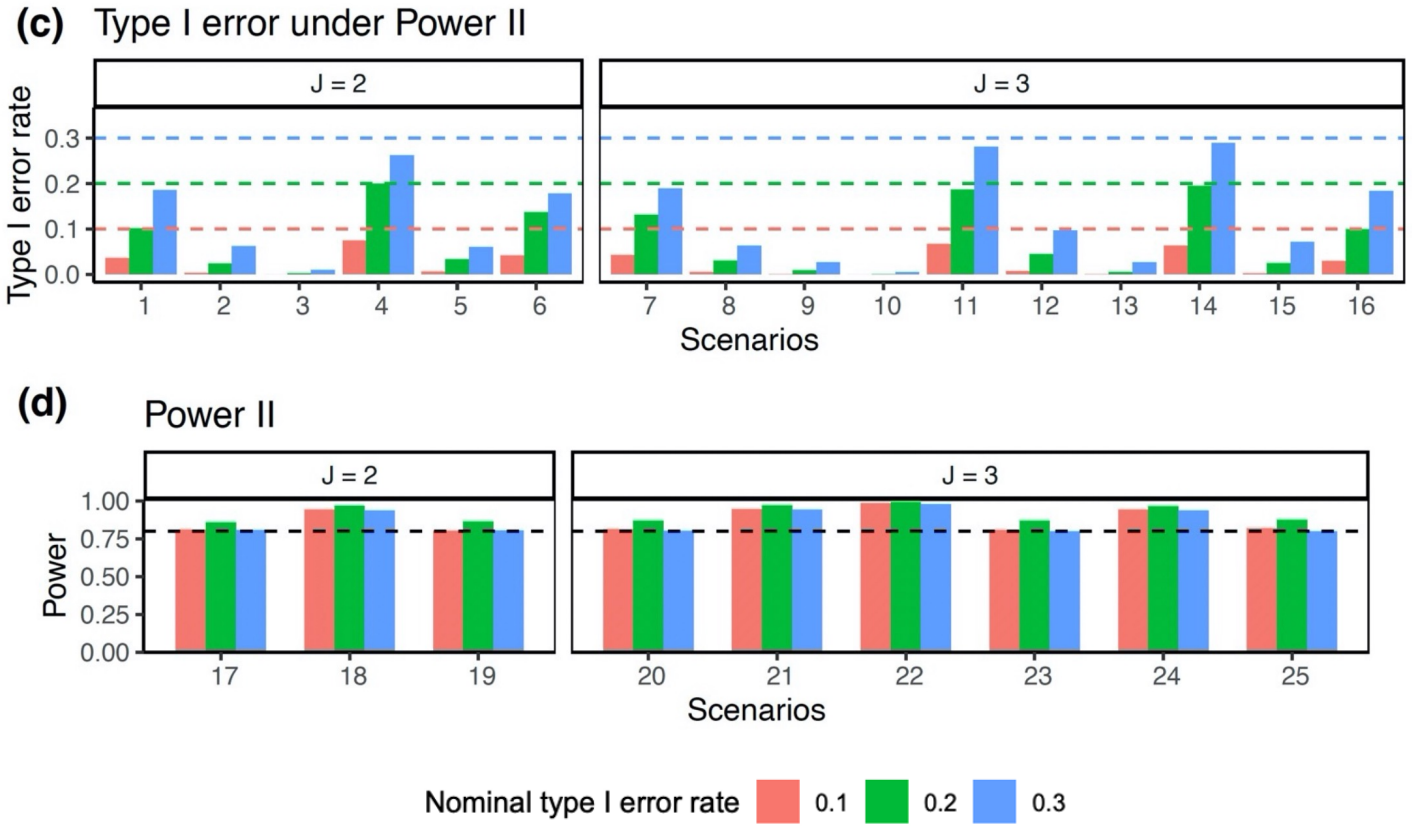
Operating Characteristics – under power I

- Type I error and power of MERIT design when $(\phi_{T,0}, \phi_{T,1}) = (0.4, 0.2)$, and $(\phi_{E,0}, \phi_{E,1}) = (0.2, 0.4)$.



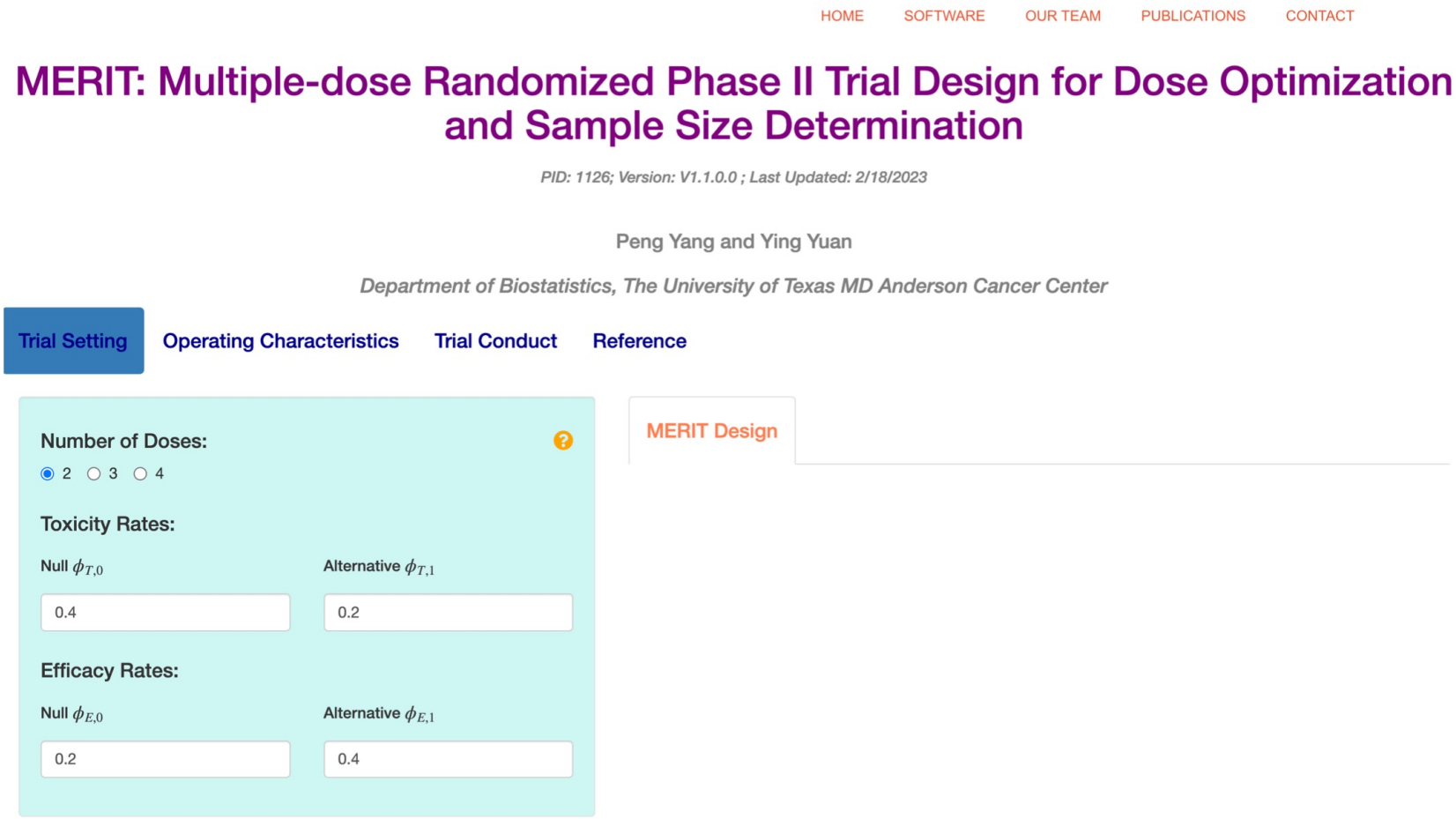
Operating Characteristics – under power II

- Type I error and power of MERIT design when $(\phi_{T,0}, \phi_{T,1}) = (0.4, 0.2)$, and $(\phi_{E,0}, \phi_{E,1}) = (0.2, 0.4)$.



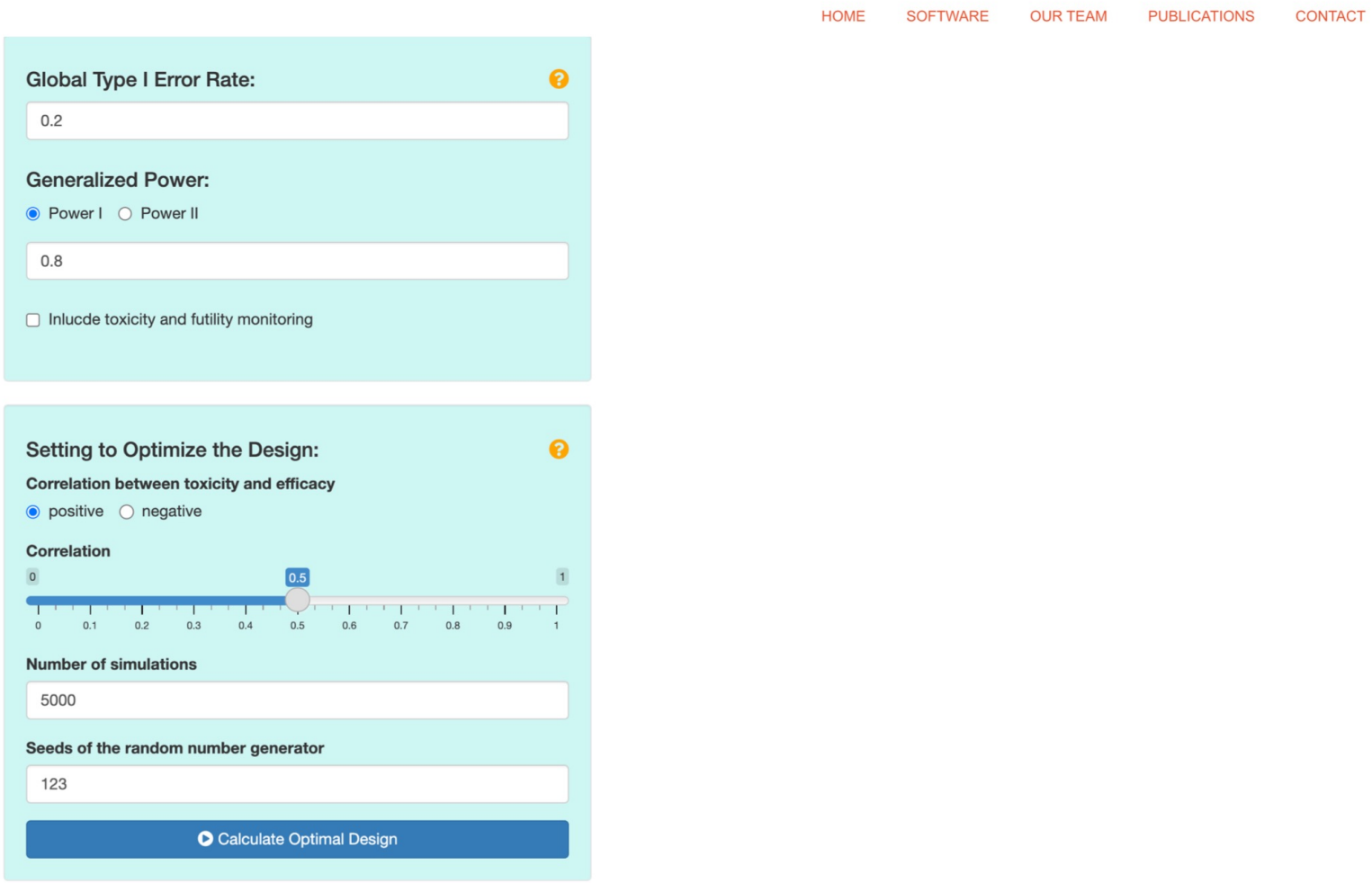
Design and Sample Size Determination for Multi-Dose Randomized Trial

↳ www.trialdesign.org



Design and Sample Size Determination for Multi-Dose Randomized Trial

www.trialdesign.org

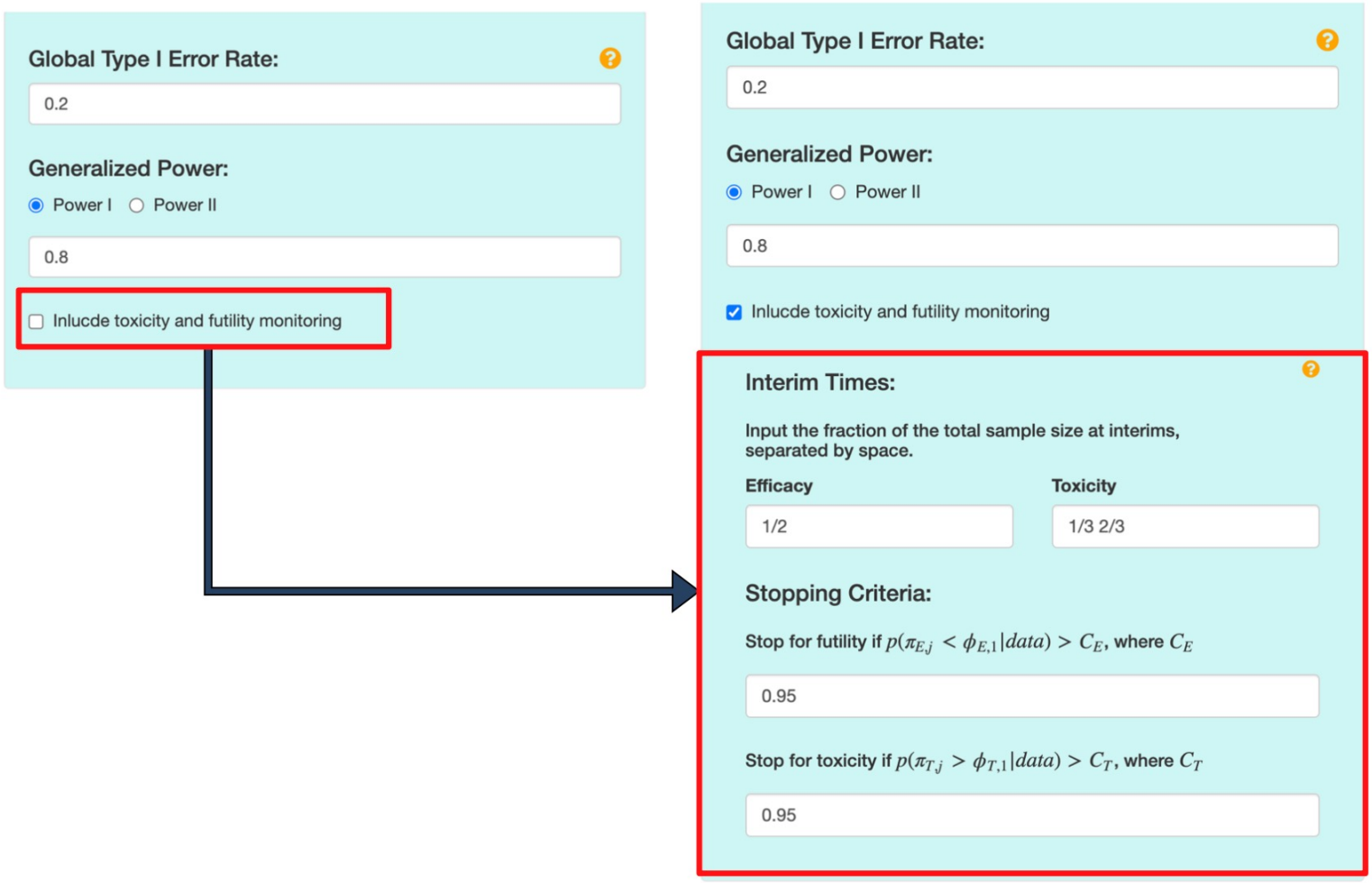


Design and Sample Size Determination for Multi-Dose Randomized Trial

↳ www.trialdesign.org

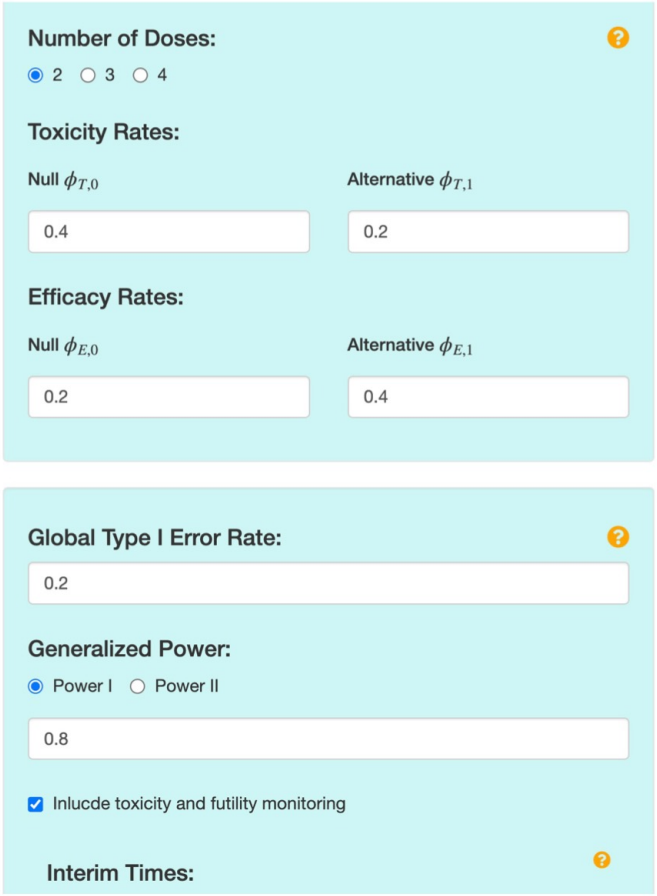
Design and Sample Size Determination for Multi-Dose Randomized Trial

↳ www.trialdesign.org



Design and Sample Size Determination for Multi-Dose Randomized Trial

↳ www.trialdesign.org



MERIT Design

Download MERIT Design

Design Description

In this trial, the toxicity rates of 0.4 and 0.2 are considered unacceptable and acceptable, respectively, while the efficacy rates of 0.2 and 0.4 are considered unacceptable and acceptable, respectively. In order to control the global Type I error rate at 0.2 and achieve a global generalized power I of 0.8, a minimum sample size of 45 per arm is required. The generalized power I is defined as the probability of rejecting the null hypothesis that none of the doses are considered optimal biological doses (OBD) admissible and all doses identified as OBD admissible are truly safe and efficacious given the alternative hypothesis that at least one dose is OBD admissible. At the end of the trial, perform isotonic regression on toxicity and efficacy data across all doses. A dose will be considered OBD admissible if the isotonically transformed number of toxicity <= 13 and the isotonically transformed number of efficacy >= 13.

During the trial, the toxicity and efficacy of each dose arm will be monitored independently using the stopping criteria outlined in Table 1. If the isotonically transformed toxicity and efficacy acrosss topping boundaries, enrollment in that particular dose arm will be suspended.

Table 1. Stopping boundaries for toxicity and efficacy.

CSV Excel PDF Print Search:

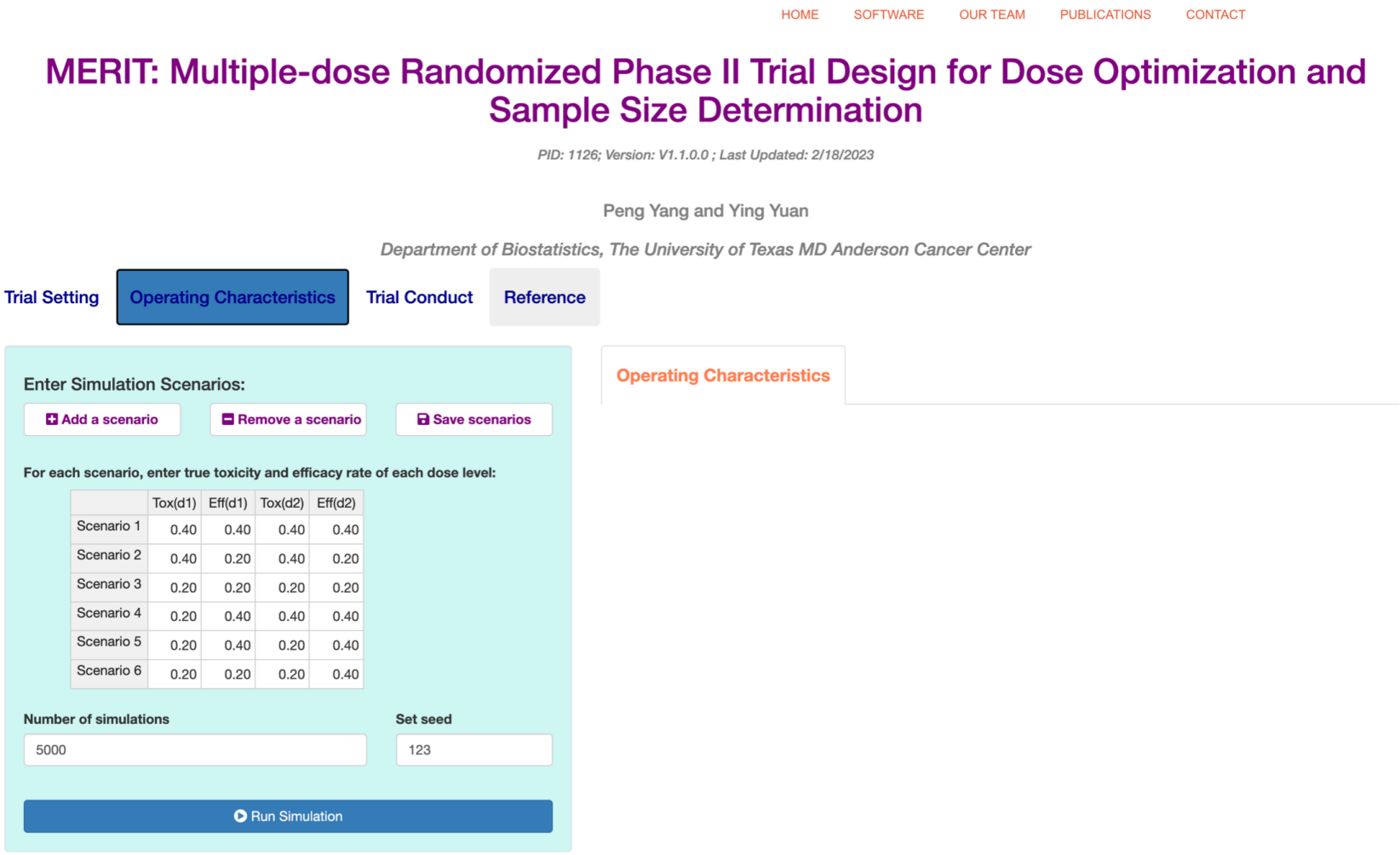
# of patients treated	Stop if # toxicity >=	Stop if # efficacy <=
15	6	NA
23	NA	5
30	10	NA

Showing 1 to 3 of 3 entries Previous 1 Next

Note: 'NA' means that this endpoint will not be used to make go/no-go decision at the interim

Design and Sample Size Determination for Multi-Dose Randomized Trial

↳ www.trialdesign.org



Design and Sample Size Determination for Multi-Dose Randomized Trial

www.trialdesign.org

Operating Characteristics

Copy | CSV | Excel | Print

Search:

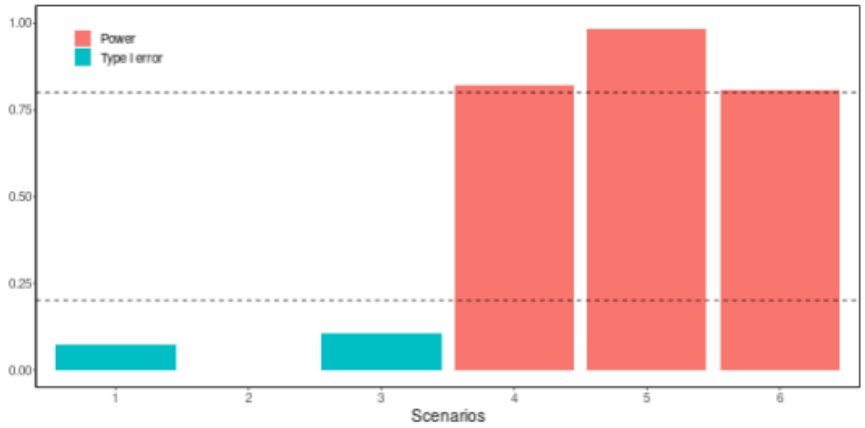
Scenarios	Metrics	Values	Average sample size
1	Type I error	0.075	45
2	Type I error	0.000	45
3	Type I error	0.104	45
4	Power	0.821	45
5	Power	0.982	45
6	Power	0.805	45

Showing 1 to 6 of 6 entries

Previous 1 Next

Download Figure 1

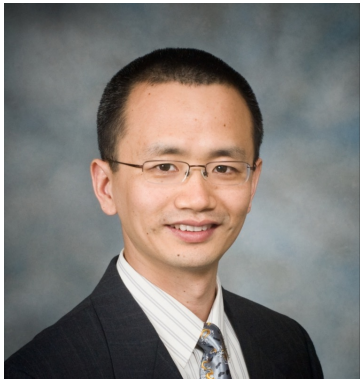
Figure 1. Type I error and power of MERIT design when unacceptable and acceptable toxicity rates are 0.4 and 0.2, and unacceptable and acceptable efficacy rates are 0.2 and 0.4. The horizontal dashed lines represent the nominal values of type I error (0.2) and power I (0.8).



Design and Sample Size Determination for Multi-Dose Randomized Trial

Discussion

- In this project, we proposed a MERIT design for dose optimization by controlling generalized type I error and power while optimizing the sample size.
- The implementation of MERIT is simple, which only involve comparison of the observed number of toxicity and efficacy with prespecified decision boundaries.
- We have developed R shiny app on www.trialdesign.org to facilitate the application of MERIT design.



THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

Design and Sample Size Determination for Multi-Dose Randomized Trial

Reference

- Yang, P., Li, D., Lin, R., Huang, B., Yuan, Y. (2023) Design and Sample Size Determination for Multiple- dose Randomized Phase II Trials for Dose Optimization, https://doi.org/10.48550/arXiv.2302.09612
- Thall, P., Simon, R. & Ellenberg, S. (1988) Two-stage selection and testing designs for comparative clinical trials. Biometrika. 75, 303-310.
- Thall, P., Simon, R. & Ellenberg, S. (1989) A two-stage design for choosing among several experimental treatments and a control in clinical trials. Biometrics. pp. 537-547.
- U.S. Food And Drug Administration. (2022) Project Optimus: Reforming the dose optimization and dose selection paradigm in oncology. Silver Spring, MD: Food and Drug Administration.
- U.S. Food And Drug Administration. (2023) Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases. Rockville, MD: Food and Drug Administration.
- Whitehead, J. (1986) Sample sizes for phase II and phase III clinical trials: an integrated approach. Statistics In Medicine. 5, 459-464.

Thank you