## Contents

### I Phase Ia Single-Agent Dose-Finding Designs

1 Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

1.1 Introduction ......................................................... 3
1.2 User Interface and Tutorial ................................. 5
  1.2.1 Overview ...................................................... 5
  1.2.2 Simulation Setup ........................................... 7
  1.2.3 Simulation Results ......................................... 16
  1.2.4 Decision Table ............................................... 27
  1.2.5 MTD Estimation ............................................... 30
1.3 Statistical Methods Review ................................. 32
  1.3.1 The 3+3 Design ............................................... 32
  1.3.2 The Continuous Reassessment Method (CRM) ............ 34
  1.3.3 The Bayesian Logistic Regression Method (BLRM) ......... 37
  1.3.4 The Modified Toxicity Probability Interval (mTPI) Design . 42
  1.3.5 The Modified Toxicity Probability Interval-2 (mTPI-2) Design .... 47
  1.3.6 The Modified Cumulative Cohort Design (mCCD) .......... 52
  1.3.7 The i3+3 Design .............................................. 58

2 Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment .... 62

2.1 Introduction ......................................................... 62
2.2 User Interface and Tutorial ................................. 64
  2.2.1 Overview ...................................................... 64
  2.2.2 Simulation Setup ........................................... 66
II Phase Ib Expansion Cohort Designs

5 Multiple Cohort Expansion

5.1 Introduction ............................................. 200

5.2 User Interface and Tutorial .............................. 201

5.2.1 Overview .............................................. 201

5.2.2 Case Study ............................................ 202

5.2.3 Quick Demo ........................................... 203

5.2.4 Data Analysis ........................................ 206

5.3 Statistical Methods Review ............................ 208

5.3.1 Multiple Cohort Expansion (MUCE) Method .... 208

III Phase II Single-Arm Continuous Monitoring

6 Bayesian Efficacy Monitoring with Predictive Probability

6.1 Bayesian Efficacy Monitoring via Predictive Probability ........................................... 213

6.1.1 Model .................................................. 213

6.1.2 Decision Criteria .................................... 214

6.1.3 Design ............................................... 215

6.1.4 An Example .......................................... 215

7 Bayesian Efficacy Monitoring with Posterior Probability

7.1 Bayesian Efficacy Monitoring via Posterior Probability ............................................. 217

7.1.1 Model .................................................. 217

7.1.2 Decision Criteria .................................... 217

7.1.3 Design ............................................... 218

7.1.4 An Example .......................................... 219
VI Sample Size Calculation

12 Sample Size Calculation for Binary Outcome

12.1 Single arm

12.1.1 Test Objective: Equality

12.1.2 Test Objective: Equivalence

12.1.3 Test Objective: Non-Inferiority/Superiority

12.1.4 Cohen’s Kappa

12.2 Two arms (independent)

12.2.1 Test Objective: Equality

12.2.2 Test Objective: Equivalence

12.2.3 Test Objective: Non-Inferiority/Superiority

12.3 Two arms (paired): McNemar’s Test

12.3.1 Methods

12.3.2 Input and Output

12.3.3 An Example (Two-arms (paired) McNemar’s Test)

13 Sample Size Calculation for Continuous Outcome

13.1 Single arm

13.1.1 Test Objective: Equality

13.1.2 Test Objective: Non-Inferiority/Superiority

13.1.3 Test Objective: Equivalence

13.1.4 Test Objective: Correlation

13.2 Two arms (independent)

13.2.1 Test Objective: Equality

13.2.2 Test Objective: Equivalence

13.2.3 Test Objective: Non-Inferiority/Superiority

13.3 Two arms (paired)

13.3.1 Methods

13.3.2 Input and Output

13.3.3 An Example (Two-arms (paired) Equality Test)

13.4 Multiple arms

13.4.1 Methods

13.4.2 Input and Output
13.4.3 An Example (Multiple-arms One-Way ANOVA Test) 310

14 Sample Size Calculation for Time-to-Event Outcome 312
  14.1 Single arm 313
    14.1.1 Methods 313
    14.1.2 Input and Output 314
    14.1.3 An Example (Single-arm One-sided Test) 314
  14.2 Two arms 316
    14.2.1 Methods 316
    14.2.2 Input and Output 317
    14.2.3 An Example (Two-arms One-sided Test) 317

15 Simon’s Two-Stage Design 320
  15.1 Method 320
  15.2 Program Input and Output 321
  15.3 Protocol Template 322

Reference 324
Part I

Phase Ia Single-Agent Dose-Finding Designs
1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

1.1 Introduction

This module is about design and conduct of cohort-based phase I dose-finding trials for a single agent. The term “cohort-based” here means that patients are enrolled in cohorts, and dose-escalation decisions are also made in cohorts.

The primary objective of phase I trials is to identify the maximum tolerated dose (MTD), defined as the highest dose with a DLT rate less than or close to a prespecified targeted rate $p_T$ (say, $p_T = 1/6$ or $1/3$). During the past three decades, a large number of designs have been developed for phase I trials. Figure 1.1 lists 12 representative designs over time. The 3+3 design by Storer (1989) has been the most popular design among physicians due to its simplicity in practice. It is a rule-based design and adaptively moves up and down cross doses by assigning three patients per cohort until the MTD is identified. Disadvantages of 3+3 are mainly the lack of reliability to identify the correct MTD (Chen et al., 2009), the lack of flexibility to accommodate patients drop-out or over-enrollment, and the poor statistical operating characteristics in terms of safety and reliability (Ji and Wang, 2013; Nie et al., 2016). Since 1990, many new methods, especially Bayesian methods, have been developed to guide dose escalation. The continual reassessment method (CRM) is the first Bayesian model-based design proposed by O’Quigley et al. (1990). It uses information from all doses to guide decision making. Neuenschwander et al. (2008) extend the CRM and propose the Bayesian logistic regression model (BLRM). Both CRM and BLRM use parametric dose-response curves for statistical modeling and inference. Founded on sound statistical principles, both designs exhibit superior performance when compared with 3+3. However, they are complex and need strong statistical input to safeguard the practical deployment, which makes them challenging for clinicians to comprehend and implement in practice. In the recent decades,
the landscape of phase I dose-finding designs has been rapidly shifting, noticeably marked by the emergence of interval-based designs, such as the toxicity probability interval (TPI) design (Ji et al., 2007) and two subsequent modifications, the mTPI (Ji et al., 2010; Ji and Wang, 2013) and mTPI-2 (Guo et al., 2017) designs. In parallel, the cumulative cohort design (CCD) (Ivanova et al., 2007) and the Bayesian optimal interval (BOIN) design (Liu and Yuan, 2015a) further simplify the statistical inference based on a point estimate of toxicity probability and prespecified interval boundaries. BOIN is an overly refined version of CCD, in which the interval boundaries are generated based on an ad-hoc objective function that creates theoretically shaky results. In our U-Design platform, we decide to adopt and modify the CCD design, rather than BOIN, following our principle to promote sound methodologies. Finally, in 2019, the evolutionary step of phase I dose-finding designs spirals back to the rule-based approaches in the form of the i3+3 design (Liu et al., 2020), which shows the potential of smart rule-based designs that can achieve comparable operating characteristics to model-based designs.

In this module of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment, U-Design performs trial simulation to examine the operating characteristics of seven designs, including i3+3 (Liu et al., 2020), mTPI-2 (Guo et al., 2017), CRM (O’Quigley et al., 1990), 3+3 (Storer, 1989), mTPI (Ji et al., 2010), modified CCD (mCCD) (Ivanova et al., 2007) and BLRM (Neuenschwander et al., 2008) designs. Also, the decision table generation and the MTD estimation are incorporated in this module, so that users may generate the decision tables to guide trial conduct and estimate the MTD after trial completion. §1.2 introduces the user interface and tutorial of launching trial simulations and examining results, as well as generating decision tables and estimating MTD. A statistical review of all seven designs are provided in §1.3.
1.2 User Interface and Tutorial

1.2.1 Overview

Entering the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment page, users will see four main tabs: Simulation Setup, Simulation Results, Decision Table and MTD Estimation. The first two tabs allow users to conduct simulations and visualize/download simulation results, and the next two tabs allow users to generate decision tables and estimate the MTD, respectively. In the Simulation Setup tab, there are three steps (Figure 1.2): 1) Set trial parameters, 2) Select designs, and 3) Generate scenarios. Users need to complete the steps 1-3 to set up simulations for a single design or multiple designs. Upon completing steps 1-3, users click the “Launch Simulation” button at the bottom of the page. Users may also click the “Reset” button next to Launch Simulation to clear all settings. After the simulation is launched, the results of simulations will be displayed in the Simulation Results tab. The simulation process can be monitored in real time at the top of the Simulation Results tab. Detailed steps of using this module are elaborated next in §1.2.2-§1.2.5.
Figure 1.2: Simulation Setup in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
1.2.2 Simulation Setup

In the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module, U-Design provides seven designs, i3+3, mTPI-2, CRM, 3+3, mTPI, mCCD and BLRM, for simulation. Users can choose up to design configurations for simultaneous comparison in the Simulation Setup tab each time. A design configuration means a design such as i3+3, along with the designs settings, such as sample size. Request to allow more than four design configurations by emailing admin@laiyaconsulting.com.

1.2.2.1 Step 1: Set trial parameters

Specify the target toxicity probability ($p_T$), number of simulations ($n_{\text{sim}}$), and random seed of simulation ($R_{\text{seed}}$) for the simulated trials. See Figure 1.3. Hover mouse over each trial parameter, and a description will be displayed explaining the meaning of the parameter. The detailed explanation of the above three input arguments is provided in Table 1.1.

![Figure 1.3: Set trial parameters in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.]

1.2.2.2 Step 2: Select designs

To select a design, click the button with the design’s name on it. Up to four design configurations may be selected for comparison.

Click the “More” link to expand the design list to see all the seven designs and click the “Less” to collapse the list.

Check the “Apply Stopping Rule” box to apply an ad-hoc stopping rule of reaching the maximum number of patients at a dose level during the trial conduct. See the detailed rules in Table 1.2 and §1.3.

Click the “Delete” button to remove the selected designs.

Design parameters can be modified in the input box of corresponding row. Hover mouse over each design parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 1.2.
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

Table 1.1: Input parameters for trials in the **Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment** module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T$</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). The main objective of phase I clinical trials is to find the highest dose with a toxicity probability closest to or lower than $p_T$. Default value is 0.3.</td>
</tr>
<tr>
<td>$n_{sim}$</td>
<td>The number of simulated trials</td>
<td>The maximum number of simulated trials allowed is 10,000. Default value is 1,000.</td>
</tr>
<tr>
<td>$R_{seed}$</td>
<td>The random seed of simulation</td>
<td>A random seed is a number used to initialize a pseudorandom number generator in the simulation. Default value is 32432.</td>
</tr>
</tbody>
</table>

Figure 1.4: Select designs in the **Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment** module.
1.2. User Interface and Tutorial

1.2.2. Simulation Setup

Table 1.2: Input parameters for designs in the **Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment** module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$ (all designs)</td>
<td>Sample size</td>
<td>The maximum number of patients to be treated in the trial. The upper limit is set at 100 since the number of patients that are enrolled in phase I clinical trial is typically small. Default value is 30.</td>
</tr>
<tr>
<td>$d_{\text{start}}$ (all designs)</td>
<td>Starting dose level</td>
<td>The starting dose level in the simulated trials. Default value is 1.</td>
</tr>
<tr>
<td>$n_{\text{cohort}}$ (except 3+3)</td>
<td>Cohort size</td>
<td>The number of patients in each cohort. Default value is 3.</td>
</tr>
<tr>
<td>$\epsilon_1, \epsilon_2$ (i3+3, mTPI, mTPI-2, mCCD, BLRM)</td>
<td>$\epsilon_1$: lower margin $\epsilon_2$: higher margin</td>
<td>Two small fractions used to define the equivalence/target interval of the MTD. Any dose with a toxicity probability falling into the interval $[p_T - \epsilon_1, p_T + \epsilon_2]$ is considered an acceptable dose MTD. Default values for both are 0.05.</td>
</tr>
<tr>
<td>$p_{\text{EWOC}}$ (BLRM)</td>
<td>Cutoff probability of escalation with overdose control</td>
<td>The threshold of controlling the probability of excessive or unacceptable toxicity. Default value is 0.25.</td>
</tr>
<tr>
<td>$\delta$ (CRM)</td>
<td>Half-width</td>
<td>The halfwidth of the indifference interval in selecting the skeleton of the model. Default value is 0.05.</td>
</tr>
<tr>
<td>$K$ (except 3+3)</td>
<td>Maximum number of patients at a dose level</td>
<td>A number used in the “Stopping Rule” that stops a trial if 1) the dose-assignment decision is to escalate to the next higher dose and there has been $K$ patients enrolled at that dose; or 2) the dose-assignment decision is to stay at the current dose and there has been $K$ patients enrolled at that dose; or 3) if the dose-assignment decision is to de-escalate to the previous lower dose and there has been $K$ patients enrolled at that dose; Default value is 12.</td>
</tr>
</tbody>
</table>
1.2.2.3 Step 3: Generate scenarios

There are two ways to generate scenarios, automatically (in below Auto Generation tab, see Figure 1.5) or through manual construction (in below Manual Construction tab, see Figure 1.6). Once scenarios are generated, click the “Launch Simulation” button at the bottom of the page to run \( n_{\text{sim}} \) (set in step 1) simulations, for each scenario and selected design (set in step 2) combination, assuming \( p_T \) (set in step 1).

**Auto Generation** (Figure 1.5)

Select the number of doses \( n_{\text{dose}} \) (\(3 \leq n_{\text{dose}} \leq 10\)) from the dropdown box. Upon selection, five or six scenarios will be created automatically, each of which contains the true toxicity probabilities for \( n_{\text{dose}} \) dose levels. These generated scenarios are displayed and editable. The detailed algorithm for scenarios auto generation is provided next.

**Manual Construction** (Figure 1.6)

Follow the instructions below to manually construct scenarios. Then click the “Add” button to create these scenarios. The format of input must comply with the following instructions.

- Scenarios should be separated by line breaks;
- The parameters in one scenario should be ordered in accordance with this sequence:
  
  Target toxicity probability, Number of simulated trials, True toxicity probabilities of all the dose levels;
- Each parameter must be separated by a white space or comma.

For example, by inputting “0.2 1000 0.05 0.1 0.15 0.2” or “0.2,1000,0.05,0.1,0.15,0.2”, a scenario is presented that runs 1000 simulated trials with a target \( p_T \) = 0.2 and true toxicity probabilities of four dose levels, 0.05, 0.1, 0.15 and 0.2.

The generated scenarios are displayed as a list (Figures 1.5 and 1.6) which appears below the generation section. An interactive chart will also be generated to visually display the shape of true toxicity probabilities for each scenario.
Figure 1.5: Automatically generate scenarios in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

Figure 1.6: Manually generate scenarios in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
Algorithm for Auto Generation

By entering the number of candidate dose levels \( n_{dose} \), five or six scenarios are generated automatically. See Figure 2.7 for an illustration. They represent the four types of dose-response shapes below.

<table>
<thead>
<tr>
<th>Types</th>
<th>Dose-Response Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal</td>
<td>Some doses are tolerable but some are overly toxic, AND there exists at least one dose level close to the target ( p_T ) or falling within the equivalence interval ([p_T - \epsilon_1, p_T + \epsilon_2]).</td>
</tr>
<tr>
<td>Safe</td>
<td>All doses are safe and tolerable with the true toxicity probabilities smaller than the target ( p_T ) or the lower boundary of equivalence interval ((p_T - \epsilon_1)).</td>
</tr>
<tr>
<td>Toxic</td>
<td>All doses are overly toxic with the true toxicity probabilities larger than the target ( p_T ) or the upper boundary of equivalence interval ((p_T + \epsilon_2)).</td>
</tr>
<tr>
<td>Steep</td>
<td>Some doses are tolerable but some are overly toxic, AND there is a steep jump in the toxicity probability between two adjacent doses, AND there is no dose close to the target ( p_T ) or falling within the equivalence interval ([p_T - \epsilon_1, p_T + \epsilon_2]).</td>
</tr>
</tbody>
</table>

Two “Steep” scenarios are generated, with the toxicity probability steep jump occurring at the first or second half of the doses.

Similarly, two “Ideal” scenarios might be generated, with the MTD placed in the first or second half of the doses.
Figure 1.7: An example of four main types of scenarios. Five dose levels are considered for escalation. The target toxicity probability is $p_T = 0.25$, and the equivalence interval is EI=[0.2, 0.3]. Six different lines represent four main types of scenario, respectively. In “Ideal” scenarios (Lines 1 and 2), doses 2 and 4 are the true MTD with toxicity probability falling within the EI, respectively. In “Safe” scenario (Line 3), all doses are safe with toxicity probabilities smaller than the target $p_T = 0.25$. “Toxic” scenario (Line 4) gives a contrary situation to the “Safe” scenario, where all doses are overly toxic with the toxicity probabilities larger than the target $p_T = 0.25$. The remaining two lines (Lines 5 and 6) are “Steep” scenario, in which some doses are tolerable but some are overly toxic and there is a steep jump in the toxicity probability occurring at the first or second half of the doses (between doses 4 and 5 in Line 5, and between doses 1 and 2 in Line 6).
1.2. User Interface and Tutorial
1.2.2. Simulation Setup

1.2.2.4 Launch Simulation

Once the steps 1-3 are completed, users can conduct simulated clinical trials to examine the operating characteristics of the selected designs using the selected scenarios, by clicking the “Launch Simulation” button at the bottom of Simulation Setup tab (Figures 1.5 and 1.6). A green “Launch Successful” message will be displayed on the website as in Figure 1.8 to indicate that the simulation has been successfully launched. Users may click the “Proceed To Simulation Results” button in the pop-up box to track the simulation processing status and simulation results.

![Launch Successful Message](image)

Figure 1.8: “Launch Successful” message after launching simulation in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
1.2.3 Simulation Results

In the Simulation Results tab, users can view and delete the simulation progress and simulation results (§1.2.3.1), inspect the escalation process in two simulated trials (§1.2.3.2), restore the simulation settings if needed (§1.2.3.3), and download intelligent simulation reports (§1.2.3.4). Specifically, all the simulation results (figures and tables) can be downloaded in Word format, accompanying the statistical sections in a trial protocol. Hereinafter, we use simulation results and operating characteristics interchangeably.

1.2.3.1 View simulation results

In the Simulation Results tab, the Running Simulations panel exhibits the progress of ongoing simulation (Figure 1.9). The ongoing simulations are displayed in ascending order by the launch time. Click the icon “×” to delete the corresponding simulation.

![Figure 1.9: Simulation progress in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.](image)

Once the simulations are completed, the Running Simulations panel in Figure 1.9 will disappear, green “simulation result created” massages will appear instead and stay at the same place of the Running Simulations panel unless explicitly dismissed by clicking the icon “×” at the end of the corresponding row, and the simulation results will be automatically loaded into the Simulation History panel (Figure 1.10), with the blue bold Unread shown to indicate new results. All the previously completed simulations are also listed in the Simulation History panel. Simulation results for other modules can also be viewed under the Simulation History by dropping down the “Select a module” button (Figure 1.10). Click the button to delete the selected simulation results.
1.2. User Interface and Tutorial
1.2.3. Simulation Results

Figure 1.10: Simulation Results in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

Click the button to unfold the simulation results (Figure 1.11). The design settings are firstly displayed at the top of each simulation study (Figure 1.11). Then the results of simulation are shown as plots and tables below.

Figure 1.11: View the simulation results in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
## Simulation Result Plots

There are four sections in the Simulation Result Plots:

A. Line plots showing four summary statistics of the simulation results for all the designs (Figure 1.12), including **Prob. of Selecting MTD**, **Prob. of Toxicity**, **Prob. of Selecting Does-over-MTD**, and **Prob. of No Selection**.

B. A table of mean and standard deviation (s.d.) for the four summary statistics (Figure 1.13).

C. [Optional] An empirical CRM decision table if CRM is selected in the simulation (Figure 1.14).

D. [Optional] An empirical BLRM decision table if BLRM is selected in the simulation (Figure 1.15).

### A. Line plots:

- The four summary statistics are part of operating characteristics of the designs. They are explained in full detail next.
  - **Prob. of Selecting MTD**: The probability of selecting the true MTD, defined as the proportion of simulated trials that correctly select the true MTD. The higher the value, the better the design.
    - For interval-based designs (i3+3, mTPI, mTPI-2, BLRM, & mCCD), the true MTDs are defined as the dose levels of which the true toxicity probabilities fall into the equivalence interval \([p_T - \epsilon_1, p_T + \epsilon_2]\); if none of the doses have a toxicity probability that falls into the equivalence interval, the true MTD is defined as the dose with the highest toxicity probability below \(p_T\). For the non-interval-based designs, 3+3 and CRM, the true MTDs is defined as the dose levels with the highest toxicity probabilities lower than or equal to \(p_T\).
    - To compare the operating characteristics of multiple designs submitted in a simulation study, the definition of MTD should be unified. If any of interval-based designs (i3+3, mTPI, mTPI-2, BLRM, & mCCD) are used in the simulation, the dose levels of which the true toxicity probabilities fall into the widest equivalence interval \([p_T - \max\{\epsilon_1\}, p_T + \max\{\epsilon_2\}]\) are defined as the true MTDs. Here, \(\max\{\cdot\}\) is taken over the designs. If none of the doses fall in, the dose with the highest toxicity probability that is below \(p_T\) is the true MTD. For example, consider a case in which users compare four designs, mTPI, mTPI-2, CRM and 3+3, in a simulation study targeting \(p_T = 0.3\). Suppose \(\epsilon_1 = 0.02\) and \(\epsilon_2 = 0.05\) for mTPI, and \(\epsilon_1 = 0.05\) and \(\epsilon_2 = 0.03\) for mTPI-2. In this case, the true MTD is the dose levels with toxicity...
probabilities in [0.3-0.05, 0.3+0.05]; if none of the doses have a toxicity probability in [0.3-0.05, 0.3+0.05], the dose with the highest toxicity probability lower than 0.3 is the true MTD.

* If a scenario does not have any MTD (e.g., all doses have toxicity probabilities larger than the target $p_T$), no selection is the right decision. In this case, the probability of selecting the true MTD is the probability of no selection.

- **Prob. of Toxicity**: The proportion of patients who have experienced DLT across all the simulated trials. The lower the number, the fewer patients having DLTs under the design.

- **Prob. of Selecting Does-over-MTD**: The probability of selecting the dose levels above the true MTD, which is defined by the proportion of simulated trials that select a dose higher than the true MTD at the end of the trial. The lower the value, the better the safety of the design.

- **Prob. of Overdosing Allocation**: The average proportion of patients who are assigned to doses higher than the MTD by the design across all the simulated trials.

- For each line plot, the x-axis is the index of scenario and the y-axis is the value of summary statistics. Lines with different colors represent different designs.

- The plots are interactive for better visualization.
  - Hover the mouse on a dot and a box will display the value of each design at the corresponding scenario (e.g. top left plot in Figure 1.12: Prob. of Selecting MTD).
  - Hover the mouse on the design label to highlight the corresponding line and fade the others (e.g. bottom right plot in Figure 1.12: Prob. of Overdosing Allocation).
  - Click the design label to hide the corresponding line and click again to change it back (e.g. top right plot in Figure 1.12: Prob. of Toxicity).
Figure 1.12: Simulation result plots in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

B. Simulation summary table: Figure 1.13 shows the mean±sd of the summary statistics across all scenarios for each design.

<table>
<thead>
<tr>
<th>Summary of Performance</th>
<th>Design 1 (3&gt;3)</th>
<th>Design 2 (CRM)</th>
<th>Design 3 (BLRM)</th>
<th>Design 4 (3&gt;3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob. of Selecting MTD</td>
<td>0.434 ± 0.11</td>
<td>0.443 ± 0.096</td>
<td>0.182 ± 0.144</td>
<td>0.241 ± 0.083</td>
</tr>
<tr>
<td>Prob. of Toxicity</td>
<td>0.229 ± 0.072</td>
<td>0.242 ± 0.078</td>
<td>0.189 ± 0.085</td>
<td>0.215 ± 0.08</td>
</tr>
<tr>
<td>Prob. of Selecting Dose-over-MTD</td>
<td>0.178 ± 0.129</td>
<td>0.213 ± 0.152</td>
<td>0.043 ± 0.047</td>
<td>0.066 ± 0.059</td>
</tr>
<tr>
<td>Prob. of Overdosing Allocation</td>
<td>0.167 ± 0.158</td>
<td>0.185 ± 0.17</td>
<td>0.062 ± 0.087</td>
<td>0.108 ± 0.119</td>
</tr>
</tbody>
</table>

* Mean ± Standard deviation. The statistics are calculated based on the current scenario and design setting.

Figure 1.13: Simulation summary table in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

C. CRM decision table:
An empirical CRM decision table will be provided in the simulation results if CRM is included in the simulation (Figure 1.14). This table summarizes the frequency of decisions made by CRM across all the simulated trials.
The lengths of the three colored bars in one cell represent the frequencies of the corresponding dose-finding decisions. The longer the bar, the higher the frequency. For example, the cell in the figure shows that CRM stay at the current dose 44.7% of the times when 2 out of 3 patients experience DLTs at a dose.

Figure 1.14: CRM decision table in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

D. BLRM decision table:
An empirical BLRM decision table will be provided in the simulation results if BLRM is included in the simulation (Figure 1.15). This table summarizes the frequency of decisions made by BLRM across all the simulated trials.

• The lengths of the three colored bars in one cell represent the frequencies of the corresponding dose-finding decisions. The longer the bar, the higher the frequency. For example, the cell in the figure shows that BLRM de-escalates to the previous lower doses 30.1% of the times when 1 out of 3 patients experienced DLT.

Figure 1.15: BLRM decision table in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
Simulation Result Tables

Full simulation results are presented in tabular format arranged by scenarios (Figure 1.16).

In the upper part of Figure 1.16, the first two columns summarize dose levels and their true toxicity probabilities; the remaining columns report three dose-specific summary statistics from the simulations: selection probability, average number of patients treated, and average number of toxicities (i.e. DLTs), along with their standard deviations, at each dose level. Specifically, they are

1) **Selection Prob.**: The proportion of simulated trials that select each dose level as the MTD.
2) **Average # of Patients Treated (s.d.)**: The average number of patients treated at each dose level and its standard deviation.
3) **Average # of Toxicities (s.d.)**: The average number of patients experienced DLT at each dose level and its standard deviation.

The true MTD(s) of the scenario is(are) highlighted by the orange bar. For the definition of the true MTD in the simulation results, please refer to the definition of **Prob. of Selecting MTD** in the Simulation Results Plots above (after Figure 1.11).

In the lower part of Figure 1.16, more trial-specific summary statistics are reported, mainly from five aspects: **MTD Selection**, **Patient Assignment**, **Trial Toxicity**, **Trial Stopping** and **Trial Sample Size**. Click the “More” link to show the summary statistics of **Trial Stopping** and **Trial Sample Size** and click the “Less” to collapse these results. Specifically, they are

- **MTD Selection**
  - **Prob. of Selecting MTD**: The proportion of simulated trials that select the true MTD at the end of the trial.
  - **Prob. of Selecting Does-over-MTD**: The proportion of simulated trials that select the doses higher than the true MTD at the end of the trial.
  - **Prob. of No Selection**: The proportion of simulated trials in which none of the dose levels are selected as the MTD. If a scenario does not have any MTD, this values is treated as the probability of selecting the true MTD.

For detailed descriptions, please refer to **Simulation Result Plots** section above (after Figure 1.11).

- **Patient Allocation**
  - **Prob. of Correct Allocation (s.d.)**: The average proportion of patients who are correctly assigned to the true MTD by the design across all the simulated trials and its standard deviation.
  - **Prob. of Overdosing Allocation (s.d.)**: The average proportion of patients who are
assigned to doses higher than the MTD by the design across all the simulated trials and its standard deviation.

- **Trial Toxicity**
  - **Prob. of Toxicity**: The proportion of patients experiencing DLT across all the simulated trials. For detailed descriptions, please refer to the *Simulation Result Plots* section above (after Figure 1.11).

- **Trial Stopping**
  - **Prob. of Early Stopping Trial due to Safety Rule**: The proportion of simulated trials in which the trial is stopped because the first dose level shows unacceptable toxicity.
  - **Prob. of Early Stopping Trial due to Reaching $K$**: The proportion of simulated trials in which the trial is stopped because the dose-assignment decision is to escalate/stay/de-escalate to a dose level but that dose has enrolled at least $K$ patients ($K < n$, e.g., $K = 12$).
  - **Prob. of Stopping Trial due to Reaching $n$**: The proportion of simulated trials in which the trial is stopped because the total number of patients enrolled and treated in a trial has reached or exceeded the pre-specified maximum sample size $n$.

- **Trial Sample Size**
  - **Average # of Patients Treated (s.d.)**: The average number of patients treated in the simulated trials and its standard deviation. Due to early stopping, this number is lower than or equal to $n$.

  When calculating the standard deviation, we use $n_{sim}$ as the denominator instead of $(n_{sim} - 1)$ in U-Design.
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

Figure 1.16: Simulation result tables in the **Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment** module.
1.2.3.2 Simulation trial examples

Users can visualize how a trial is conducted by clicking a “Simulation trial examples” button at the upper right corner of each simulation results table (Figure 1.16). The pop-up box (Figure 1.17) shows the dose escalation process of two simulated trials for each design.

A red or green dot indicates a patient with or without DLT, respectively. Dots within the same region of white or light blue background color indicate patients in the same cohort. The horizontal red line indicates the dose level selected as the MTD at the end of the trial. The absence of the red line indicates none of the dose levels is selected as the MTD.

![Simulation trial examples](image)

Figure 1.17: Simulation trial examples in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
1.2.3.3 **Restore simulation setup**

Users can restore the simulation settings from the simulation results by clicking the button at the upper right corner of each simulation results panel (yellow arrow in Figure 1.18). Upon clicking, the display will switch to the Simulation Setup page with the same simulation settings restored. This is useful to restore the old simulation settings for reproducible results.

![Simulation Setup](image)

**Figure 1.18:** Restore simulation setup and download simulation results in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

1.2.3.4 **Download simulation results**

There is a button at the upper right corner of each simulation results panel (green arrow in Figure 1.18). Click it to download a word file, which includes four parts:

- Part A: Complete simulation results under the designs and scenarios users added in the Simulation Setup tab;
- Part B: Intelligent template(s) for the statistical section of i3+3 and/or mTPI-2 design in a trial protocol, if users select i3+3 and/or mTPI-2 in the Simulation Setup tab;
- Part C: Detailed technical descriptions of the designs users added in the Simulation Setup tab;
- Part D: Reference

Users may select the required parts and modify them tailored for their trials or contact us via email (admin@laiyaconsulting.com) for consulting services.
1.2.4 Decision Table

In the Decision Table tab, users can generate decision tables of five designs, i3+3, mTPI, mTPI-2, mCCD and 3+3 designs, to guide the dose escalation/de-escalation during trial conduct. The CRM and BLRM designs do not provide decision tables before the trial is started. However, for both designs, U-Design provides empirical decision tables after launching simulations (§1.2.3.1).

Manually type in the maximum number of patients at a dose ($n$), target toxicity probability ($p_T$) and two small fractions ($\epsilon_1$ and $\epsilon_2$) for decision table generation (Figure 1.20). Hover mouse over each parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 1.3.

![Image](image.png)

**Figure 1.19:** Input parameters in the Decision Table tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

Click the “Generate” button to generate five decision tables for five different designs at the same time (Figure 1.20). Users can click the tabs to switch between the tables for the i3+3, mTPI-2, mCCD and 3+3 designs.

Click the “Download Decision Table” button to save the decision table of the corresponding design in word (.docx).

For each decision table, the column represents the number of patients treated at a dose, which is mostly used for the current dose, the dose currently being used to treat patients in the trial, and the row represents the number of patients among those treated at that dose who have experienced dose-limiting toxicity (DLT) events. Note that these are the counts of patients, not DLT events. For example, column 3 and row 1 means that 3 patients have been treated at the current dose and 1 of them experiences DLT. Each cell in the decision table provides the dose-assignment decision based on the readouts from the corresponding row and column. For example, for column 3 and row 1, i.e., 1 out of 3 patients experiences DLTs, the decision is “S”. The letters in the decision table represent...
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

Table 1.3: Input arguments in the Decision Table tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>Number of patients at a dose</td>
<td>The maximum number of patients to be treated at a dose. Here, the upper limit is set at 30 since the number of patients that are enrolled at a dose in phase I clinical trial is typically small.</td>
</tr>
<tr>
<td>(p_{\text{T}})</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). The main objective of phase I clinical trials is to find the highest dose with a toxicity probability closest to or lower than (p_{\text{T}}).</td>
</tr>
<tr>
<td>(\epsilon_1, \epsilon_2)</td>
<td>(\epsilon_1): lower margin, (\epsilon_2): higher margin</td>
<td>Two small fractions used to define the equivalence/target interval of the MTD. Any dose with a toxicity probability falling into the interval ([p_{\text{T}} - \epsilon_1, p_{\text{T}} + \epsilon_2]) is considered an acceptable dose MTD. Default values for both are 0.05.</td>
</tr>
</tbody>
</table>

Different dose-assignment decisions as shown below:
- “E” stands for escalating to the next higher dose,
- “S” stands for staying at the current dose,
- “D” stands for de-escalating to the previous lower dose,
- “DU” stands for de-escalating to the previous lower dose and marking the current dose and its higher doses as unacceptably toxic so that they will never be used again in the remainder of the trial.

The 3+3 decision table is fixed regardless of different trial parameters. For CRM (or BLRM), the decision table cannot be easily summarized since the dose-assignment decision under CRM (or BLRM) for a given outcome (say, 1 DLT out of 3 patients) and a given dose are random, depending on existing data in the entire trial including those at other doses. In other words, CRM (or BLRM) could stay, escalate or de-escalate when 1 out of 3 patients having DLT at a dose, which makes it impossible to provide a fixed decision table. Nevertheless, U-Design provides empirical CRM (or BLRM) decision table in the simulation section when CRM (or BLRM) is implemented in simulation trials (§1.2.3.1).
1.2. User Interface and Tutorial

1.2.4. Decision Table

Figure 1.20: Decision tables generated in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

1.2.5 MTD Estimation

In the MTD Estimation tab, users can estimate the MTD for i3+3, mTPI and mTPI-2 designs based on the isotonic regression through Pool Adjacent Violators Algorithm (PAVA), after the dose finding is completed and the DLT outcomes of all patients are collected.

Specify the target toxicity probability \( p_T \), and two small fractions to define the equivalence interval \( \epsilon_1 \) and \( \epsilon_2 \) in the design. Select the number of doses \( n_{dose} \) from the dropdown box, then an editable table will be shown below on the page (Figure 1.21). Then manually type in the observed number of toxicities (DLTs) and the number of patients treated at each dose into the table and click the “Estimate” button to estimate the MTD. Finally, the estimated MTD is highlighted in blue background as shown in Figure 1.22.

Hover mouse over each parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 1.4.

![Figure 1.21: Input parameters in the MTD Estimation tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.](image1)

![Figure 1.22: MTD estimation in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.](image2)
Table 1.4: Input parameters in the MTD Estimation tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T$</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). The main objective of phase I clinical trials is to find the highest dose with a toxicity probability closest to or lower than $p_T$.</td>
</tr>
<tr>
<td>$\epsilon_1, \epsilon_2$</td>
<td>$\epsilon_1$: lower margin $\epsilon_2$: higher margin</td>
<td>Two small fractions used to define the equivalence/target interval of the MTD. Any dose with a toxicity probability falling into the interval $[p_T - \epsilon_1, p_T + \epsilon_2]$ is considered an acceptable dose MTD. Default values for both are 0.05.</td>
</tr>
<tr>
<td>$n_{dose}$</td>
<td>The number of doses</td>
<td>The number of candidate dose levels for investigation</td>
</tr>
<tr>
<td># of DLTs</td>
<td>The number of patients with DLTs at each dose level</td>
<td>A non-negative integer number of patients with DLT at each dose level</td>
</tr>
<tr>
<td># of patients</td>
<td>The number of patients treated at each dose level</td>
<td>A positive integer number of patients treated at each dose level, which should be no less than the # of DLTs</td>
</tr>
</tbody>
</table>
1.3 Statistical Methods Review

1.3.1 The 3+3 Design

The 3+3 design (Storer, 1989) is a rule-based design which starts by allocating the first cohort of patients to the starting dose (which is often the lowest dose level) and adaptively escalates/de-escalates to the next dose level based on observed number of dose limiting toxicities (DLTs).

1.3.1.1 Design Algorithm

In 3+3, a maximum of six patients are allowed to be treated at any dose level, and the MTD is defined as the highest dose for which one or fewer DLTs occurred in six patients. Its algorithm proceeds as follows:

0. Start the trial by treating three patients at a prespecified starting dose level.

1. Escalate to the next higher dose or de-escalate to the previous lower dose according to the following rules:
   
   (a) If 0 of 3 patients has a DLT, escalate to next higher dose and treat three patients.
   
   (b) If 2 or more of 3 patients have DLTs, de-escalate to previous lower dose and treat three patients.
   
   (c) If 1 of 3 patients has a DLT, treat three more patients at current dose level.
       i. If 1 of 6 has DLT, escalate to next higher dose and treat three patients if the next higher dose has not been tried; otherwise, declare it as the MTD and stop the trial.
       
       (ii. If 2 or more of 6 have DLTs, de-escalate to previous lower dose level and treat three patients.

   (d) If the trial de-escalates to previous lower dose:
       i. If only 3 or less had been treated at the previous lower dose, treat three more patients at that dose.
       ii. If six have already been treated at the previous lower dose, stop the trial and declare the lower dose as the MTD.

2. Escalation never occurs to a dose at which two or more DLTs have already occurred.

3. If de-escalation occurs at the lowest dose, the trial is stopped.

4. Repeat steps 1-3 until either the MTD is identified or the trial is stopped for excessive toxicity.

The above algorithm can be summarized in Figure 1.23 (Yang et al., 2015).
1.3. Statistical Methods Review

1.3.1. The 3+3 Design

Figure 1.23: Schema of the 3+3 design.
1.3.2 The Continuous Reassessment Method (CRM)

CRM is a Bayesian adaptive model-based design introduced in O’Quigley et al. (1990). It assumes a parametric dose-response model in which the probability of toxicity monotonically increases with dose. The estimated dose-response curve is updated after each patient’s toxicity data is observed, and the dose closest to MTD is obtained from the updated dose toxicity curve. In the original CRM (O’Quigley et al., 1990), it is possible to escalate by more than one dose level, which may result in escalation to fairly high doses quite early. Goodman et al. (1995) proposed several practical rules for the original CRM to reduce the risk.

1.3.2.1 Probability Model

**Dose-response curve:** Denote the dose levels as $x_d$ for $d = 1, \ldots, D$, and the binary indicator of DLT for the $j$th patient as $Y_j$ for $j = 1, \ldots, n$. Let $t_j$ be the dose for patient $j$, and let $p_d = P(r(Y_j = 1|t_j = x_d)$ be the toxicity probability of dose $d$. Consider a dose-response function $p_d = \psi(x_d, \theta)$ representing the relationship between $p_d$ and $x_d$, which includes a single parameter $\theta$. Popular choice of $\psi$ includes the power model, one-parameter logistic model, and hyperbolic tangent model (Cheung, 2011). U-Design uses a simple one-parameter power model:

$$p_d = \psi(p_0, d, \theta) = p_0 \exp(\theta),$$

where $(p_{0,1}, p_{0,2}, \ldots, p_{0,D})$ are pre-specified prior toxicity probabilities (‘skeletons’), which monotonically increases with $d$. The skeletons reflect the initial guess of DLT probabilities.

**Prior specification:** Let $g(\theta)$ be the prior distribution for $\theta$, which reflects our knowledge of the dose toxicity relationship before the trial begins. In U-Design, we use the normal density $N(0, 1.16^2)$ by default (Lee and Cheung, 2011). Other choices can be gamma or exponential density.

**Estimate the probability of toxicity:** Denote the accumulated toxicity data $data \equiv \{(y_d, n_d) : d = 1, 2, \ldots, D\}$, where $n_d$ and $y_d$ are the total number of patients treated at dose $d$ and the corresponding number of patients having DLTs, respectively. Estimate the probability of toxicity $p_d$ for dose level $d$ by

$$p_d = \psi(p_0, d, E(\theta|data)), \text{ where } E(\theta|data) = \int_{-\infty}^{\infty} \theta f(\theta|data) d\theta,$$

(1.1)
1.3. Statistical Methods Review

1.3.2. The Continuous Reassessment Method (CRM)

for \(d = 1, \ldots, D\), where \(f(\theta|\text{data})\) is the posterior of \(\theta\) given by

\[
f(\theta|\text{data}) \propto \prod_{d=1}^{D} \psi(p_{0,d}, \theta)^{y_d} (1 - \psi(p_{0,d}, \theta))^{n_d-y_d} g(\theta).
\]

Calibration of the ‘skeleton’ values: Lee and Cheung (2011) proposed a fast and systematic approach for selecting the skeleton based on indifference intervals for the MTD. The approach is imbedded in U-Design by default, and users only need to specify the half-width (\(\delta\)) of the indifference interval manually to estimate the skeleton.

Specifically, assume \(\Theta = [b_1, b_{D+1}]\) is the parameter space (i.e. \(\theta \in \Theta\)) and \(H_d = [b_d, b_{d+1}]\) for \(d = 2, \ldots, D-1\) and \(H_D = [b_D, b_{D+1}]\) where \(b_d\) is the solution for \(\psi(p_{0,d-1}, b_d) + \psi(p_{0,d}, b_d) = 2p_T\) for \(d = 2, \ldots, D\). Based on Lee and Cheung (2011), define the half width of the indifference interval for the MTD (\(d\)) as

\[
\delta_d = \frac{\psi(p_{0,d+1}, b_{d+1}) - \psi(p_{0,d-1}, b_d)}{2}, \quad d = 2, \ldots, D-1.
\]

By specifying a common half-width indifference interval for all dose levels, that is \(\delta_d = \delta\), the skeletons \(p_{0,1}, \ldots, p_{0,D}\) can be obtained recursively. Given a starting dose \(\nu\), a target \(p_T\) and a prior mean of \(\theta = 0\), \(p_{0,\nu}\) can be obtained via backward substitution, i.e. \(p_T = \psi(p_{0,\nu}, 0) = p_{0,\nu}\). The remaining skeletons can be obtained by solving the following equations:

\[
\begin{align*}
\psi(p_{0,d-1}, b_d) + \psi(p_{0,d}, b_d) &= 2p_T & \text{for } d \leq \nu; \\
\psi(p_{0,d-1}, b_d) &= p_T - \delta \\
\psi(p_{0,d}, b_{d+1}) + \psi(p_{0,d+1}, b_{d+1}) &= 2p_T & \text{for } d > \nu.
\end{align*}
\]

U-Design takes \(\nu = \lceil D/2 \rceil\) as the prior guess of MTD by default. However, customized choice of \(\nu\) is provided in U-Design as well.

1.3.2.2 Design Algorithm

Dose Finding Rules: Assume patients are enrolled in cohorts. After each cohort of patients completes the DLT follow-up period, the dose to be assigned is the one that has the posterior mean probability of toxicity closest to the target \(p_T\). In other words, the next cohort of patients is assigned to dose \(d^* = \arg\min_d |\hat{p}_d - p_T|\) where \(\hat{p}_d\) is the posterior mean of toxicity probability.

Additional safety rules: In U-Design, three additional rules are applied for safety.
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

[Rule 1: Dose Exclusion] If the current dose is considered excessively toxic, i.e., \( \text{Prob}\{p_d > p_T \mid data\} > \xi \), where the threshold \( \xi \) is close to 1, say 0.95, the current and all higher doses will be excluded and never be used again in the remainder of the trial to avoid any other patients receiving treatment at those doses.

[Rule 2: Early Stop] If the current dose is the lowest dose (first dose) and is considered excessively toxic, i.e., \( \text{Prob}\{p_1 > p_T \mid data\} > \xi \), where the threshold \( \xi \) is close to 1, say 0.95, stop the trial early for safety.

[Rule 3: No-Skipping Escalation] Dose-escalation cannot increase by more than one level. That is, suppose the current dose is \( d \). If the next dose \( d^* \) satisfies \( (d^* - d) > 1 \), escalate to dose \( (d + 1) \) instead.

Here in Rules 1 and 2, \( \text{Prob}\{p_d > p_T \mid data\} \) is a function of the cumulative distribution of \( \text{beta}(\alpha_0 + y_d, \beta_0 + n_d - y_d) \). In U-Design, \( \alpha_0 = \beta_0 = 1 \) is used. Lastly, no escalation is permitted if the empirical rate of DLT for the most recent cohort is higher than \( p_T \), according to the coherence principle (Cheung, 2011).

**Trial termination:** The trial proceeds until any of the following stopping criteria is met:

1. If the prespecified maximum total sample size \( n \) is reached;
2. If the lowest dose shows excessive toxicity according to Rule 2; In this case, the trial is early stopped and the MTD cannot be determined;
3. Optional: ad-hoc rules of maximum number of patients at a dose, denoted by \( K \) (\( K < n \)):
   - If the CRM decision is “S”, to stay at the current dose, and the current dose level has enrolled \( K \) patients;
   - If the CRM decision is “E”, to escalate to the next higher dose, and the next higher dose has enrolled \( K \) patients;
   - If the CRM decision is “D”, to de-escalate to the previous lower dose, and the previous lower dose has enrolled \( K \) patients.

**MTD selection:** Once all the enrolled patients complete the DLT observation and the trial is not stopped early, the dose level \( d^{**} \) is selected as the MTD with the smallest difference of \( |\hat{p}_d - p_T| \) among all tried and safe doses \( d \), where \( \hat{p}_d \) is the posterior mean of toxicity probability for dose \( d \).
1.3.3 The Bayesian Logistic Regression Method (BLRM)

The Bayesian Logistic Regression Method (BLRM) is a model-based design proposed by Neuen-schwander et al. (2008). BLRM improves upon CRM in that it offers a more flexible representation of the dose toxicity relationship and accounts for the uncertainty associated with DLT probability point estimation during dose finding. In BLRM, one classifies the posterior probability of toxicity into four categories: under-dosing, targeted, excessive, and unacceptable toxicity, and calculates the posterior probability of DLT rate falling into four corresponding intervals at each dose. The final dose recommendation aims at maximizing the probability of targeted toxicity while controlling the probability of excessive or unacceptable toxicity at a pre-specified threshold. Besides, BLRM can also accommodate different conservatism for dose finding behavior through specification of a loss function.

1.3.3.1 Probability Model

For a set of candidate doses $d \in \{1, \ldots, D\}$, where $D$ is the number of doses, BLRM assumes a two-parameter logistic model between dose levels $x_d$ and the probability of DLT $p_d$, which is given by

$$
\text{logit}(p_d) = \log(\alpha) + \beta \log(x_d / x_{d^*}), \quad \alpha > 0, \beta > 0
$$

where $x_{d^*}$ is the reference dose, determined so that $\log(\alpha)$ is the log-odds of toxicity when $x_d = x_{d^*}$. U-Design uses a default set of doses, $x_d = 5 \times d$, and a default reference dose level $x_{d^*}$, the ceiling of $(D + 1)/2$. As a result, users do not need to input the candidate doses and reference doses manually on U-Design. However, we offer customized service allowing input of these values upon users’ requests.

1.3.3.2 Dosing Intervals and Selection

**Probability intervals:** Suppose the target probability of DLT is $p_T$ and BLRM divides the probability interval $(0, 1)$ into four categories: under-dosing $p_d \in (c_0 = 0, c_1]$, target toxicity $p_d \in (c_1, c_2]$, excessive toxicity $p_d \in (c_2, c_3]$ and unacceptable toxicity $p_d \in (c_3, c_4 = 1)$. After each patient cohort is enrolled and toxicity data are observed, the posterior distribution of $p_d$ is used to calculate the four probabilities of under-dosing, targeted, excessive and unacceptable toxicity. Based on the four probabilities, the next dose will be selected depending on one of the following two methods: minimize the Bayes risk or maximize the distance to the targeted toxicity probability subject to escalation with overdose control (EWOC).
**Method 1: Minimize the Bayes risk** A formal loss function is introduced to quantify the penalty of ending up in each of the four aforementioned intervals:

$$L(\theta, x_d) = \begin{cases} 
\ell_1 & \text{if } p_d \in (0, c_1] \\
\ell_2 & \text{if } p_d \in (c_1, c_2] \\
\ell_3 & \text{if } p_d \in (c_2, c_3] \\
\ell_4 & \text{if } p_d \in (c_3, 1) 
\end{cases}$$

Using the above loss function, one can calculate the Bayes risk $= \ell_1 \times \text{Prob}\{p_d \in (0, c_1] \mid \text{Data}\} + \ell_2 \times \text{Prob}\{p_d \in (c_1, c_2] \mid \text{Data}\} + \ell_3 \times \text{Prob}\{p_d \in (c_2, c_3] \mid \text{Data}\} + \ell_4 \times \text{Prob}\{p_d \in (c_3, 1) \mid \text{Data}\}$ and the dose minimizing the Bayes risk is selected as the next dose. In Neuenschwander et al. (2008), three different loss functions are compared in terms of dose-escalation behavior: (i) aggressive (’1-0-1-1’), (ii) conservative (’1-0-1-2’), and (iii) very conservative (’1-0-2-4’).

Depending on the compound and/or indication under study, the probability interval specification and loss function should be tailored to the specific clinical setting. However, the specification of loss function may be difficult and may complicate the interactions with clinical teams, thus the dose recommendation approach below is often used instead of the actual Bayesian decision analytic framework.

**Method 2: Maximize the distance to the target toxicity probability subject to EWOC** Babb et al. (1998) proposed to select the dose for each cohort patients as the one that maximizes the probability of targeted toxicity, i.e., $\text{Prob}\{p_d \in (c_1, c_2] \mid \text{Data}\}$ subject to the constraint that the probability of overdosing (i.e., excessive and unacceptable toxicity) does not exceed a predefined threshold $p_{EWOC}$. That is, choose the dose level subject to the constraint $\text{Prob}\{p_d \in (c_2, 1) \mid \text{Data}\} \leq p_{EWOC}$.

U-Design adopts the second method for dose recommendation by default, except that the targeted interval is defined as $(c_1 = p_T - \epsilon_1, c_2 = p_T + \epsilon_2)$ to make it consistent with settings in mTPI and mTPI-2 designs.

### 1.3.3.3 Posterior and Prior

**Prior Specification:** Model parameters $\theta = (\alpha, \beta)'$ follow a multivariate log-normal prior $\pi(\theta)$, given by

$$\log(\theta) = \begin{pmatrix} \log(\alpha) \\ \log(\beta) \end{pmatrix} \sim MVN \left\{ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma \right\}, \text{where } \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix},$$
1.3. Statistical Methods Review
1.3.3. The Bayesian Logistic Regression Method (BLRM)

where “MVN” stands for a multivariate normal distribution. Let \( \eta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \) be the hyperparameter set of the model. In U-Design we use the **quantile-based non-informative prior** calculator proposed by Neuenschwander et al. (2008) to obtain the values of \( \eta \).

The hyperparameter calculation process is based on a set of quantiles for the probabilities of toxicity that are derived from minimally informative unimodal beta distributions. Here, a beta distribution \( X \sim \text{beta}(a, b) \) is defined as a minimally informative unimodal distribution, given a prespecified quantile \( q(p) \) of the prior distribution, if (i) \( \text{Prob}\{X < q(p)\} = p \), (ii) \( a \geq 1 \) or \( b \geq 1 \) (or both), and (iii) \( a+b \) minimal. For a given prior quantile \( q(p) \), the parameters and the quantiles of a minimally informative unimodal beta distribution can be easily obtained. If \( q(p) > p \), \( \text{beta}(a, 1) \) is minimally informative unimodal if \( a = \ln(p) / \ln\{q(p)\} \). Alternatively, if \( q(p) < p \), \( \text{beta}(1, b) \) is minimally informative unimodal if \( b = \ln(1-p) / \ln\{1-q(p)\} \). Specifically, the following steps are used for this prior distribution specification:

1. Obtain the set of prior quantiles \( Q \) for the distribution of \( p_d \). In U-Design, we summarize prior information at a given dose using the median, 2.5%-th and 97.5%-th percentiles, denoted by \( q_d = \{q_d(2.5\%), q_d(50\%), q_d(97.5\%)\} \).
   (a) For the lowest dose \( d = 1 \), the prior probability of exceeding a certain threshold \( q_1(\phi_1) \) is \( \phi_1 \). In U-Design, the following default values will be used: \( \text{Prob}\{p_1 > 0.4\} = 5\% \), i.e. for the lowest dose the probability of excessive toxicity will be set to be 5 percent.
   (b) For the highest dose \( d = D \), the prior probability of falling below a certain threshold \( q_D(\phi_2) \) is \( \phi_2 \). In U-Design, the following default values will be used: \( \text{Prob}\{p_D \leq 0.2\} = 0.05 \), i.e. for the highest dose the probability of under-dosing will be set to be 5 percent.
   (c) Assuming a minimally informative unimodal beta distribution in (a) and (b) leads to prior medians for the probabilities of toxicity \( p_1 \) and \( p_D \), say \( \mu_1 = q_1(50\%) \) and \( \mu_D = q_D(50\%) \).
   (d) Prior medians \( \mu_1, \ldots, \mu_D \) are assumed to be linear in log-dose on the logit scale. This decides the minimally informative unimodal beta distributions for each dose \( d \).
   (e) For each dose \( d \), two quantiles (2.5% and 97.5%) is derived using minimally informative unimodal beta distributions with prior medians equal to \( \mu_d \).
   (f) Therefore, a set of \( D \times 3 \) quantiles are obtained, denoted by \( Q = \{q_{dk}\} \) with \( q_{dk} = q_d(\pi_k), d = 1, 2, \ldots, D, k = 1, 2, 3 \), where \( \pi_1 = 2.5\%, \pi_2 = 50\% \) and \( \pi_3 = 97.5\% \).

2. For the two-parameter logistic model the above constructed quantiles \( Q \) are then compared with the quantiles \( Q' \) coming from the bivariate normal prior distribution. We will minimize
the following criteria:

$$C(Q, Q') = \max_{d,k} |q_{dk} - q'_d|, d = 1, 2, \ldots, D, k = 1, 2, 3.$$  

The minimization of $C(Q, Q')$ leads to the optimal parameter for the prior distribution $\eta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, which can be achieved by a stochastic optimization using a Metropolis algorithm (Robert and Casella, 2013).

**Posterior Calculations:** The dose selection process described above requires the calculation of the posterior probability $\text{Prob}\{p_d \in (c_{i-1}, c_i) \mid \text{Data}\}$, for $i = 1, 2, 3, 4$, which is calculated with respect to

$$\pi(\theta \mid y, n, x) \propto e^{\sum_{d=1}^D y_d (\log(\alpha) + \beta \log(x_d/x_d^*))} \prod_{d=1}^D (1 + e^{\log(\alpha) + \beta \log(x_d/x_d^*)})^{n_d} \times \pi_0(\theta).$$

where $n = \{n_1, \ldots, n_D\}$ and $y = \{y_1, \ldots, y_D\}$ are observed toxicity data, $n_d$ and $y_d$ are the number of patients treated and having DLTs at the dose $d$, respectively. Let $\text{Data} \equiv (n, y)$, and $x = \{x_1, \ldots, x_D\}$ are candidate dose levels. Using Markov chain Monte Carlo (MCMC) simulation, the posterior inference is made based on the posterior samples drawn for $(\alpha, \beta)$ via Metropolis-Hastings algorithm.

### 1.3.3.4 Design Algorithm

**Dose Finding Rules:** Assume patients are enrolled in cohorts. After each cohort of patients completes the DLT evaluation period, the dose to be assigned by BLRM is the one that has the largest posterior probability being at the targeted interval, i.e., $\text{Prob}\{p_d \in (p_T - \epsilon_1, p_T + \epsilon_2) \mid \text{Data}\}$ subject to the constraint that the probability of overdosing does not exceed a predefined threshold $p_{\text{EWOC}}$, i.e., $\text{Prob}\{p_d \in (p_T + \epsilon_2, 1) \mid \text{Data}\} \leq p_{\text{EWOC}}$.

**Additional safety rules:** In U-Design, three additional rules are also applied for safety.

- **[Rule 1: Dose Exclusion]** If the current dose is considered excessively toxic, i.e., $\text{Prob}\{p_d > p_T \mid \text{Data}\} > \xi$, where the threshold $\xi$ is close to 1, say 0.95, the current and all higher doses will be excluded and never used again in the remainder of the trial.

- **[Rule 2: Early Stop]** If the current dose is the lowest dose (first dose) and is considered excessively toxic, i.e., $\text{Prob}\{p_1 > p_T \mid \text{Data}\} > \xi$, where the threshold $\xi$ is close to 1, say 0.95, stop the trial early for safety.

Besides, if all doses violate the EWOC rule, the trial will also be terminated early before the prespecified maximum sample size is reached.
1.3. Statistical Methods Review

1.3.3. The Bayesian Logistic Regression Method (BLRM)

-Rule 3: No-Skipping Escalation- Dose escalation cannot increase by more than one level, although dose de-escalation can (Goodman et al., 1995). That is, suppose the current dose is dose level \( d \). If the next dose \( d^* \) satisfies \( (d^* - d) > 1 \), escalate to dose \( (d + 1) \) instead.

Here in Rules 1 and 2, \( \text{Prob}\{p_d > p_T \mid \text{Data}\} \) is a function of the cumulative distribution of beta\((\alpha_0 + y_d, \beta_0 + n_d - y_d)\), and \( \alpha_0 = \beta_0 = 1 \) is used in U-Design by default, where \( y_d \) and \( n_d \) are the number of patients treated and the number of DLTs at the dose \( d \).

**Trial termination:** The trial proceeds until any of the following stopping criteria is met:

1. If the prespecified maximum total sample size is reached;
2. If the lowest dose shows excessive toxicity according to Rule 2; In this case, the trial is stopped early and the MTD cannot be determined;
3. Optional: ad-hoc rules of maximum number of patients at a dose, denoted by \( K \) \((K < n)\):
   - If the BLRM decision is “S”, to stay at the current dose, and the current dose has enrolled \( K \) patients;
   - If the BLRM decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled \( K \) patients;
   - If the BLRM decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled \( K \) patients.

**MTD selection:** Once all the enrolled patients complete the DLT observation and the trial is not stopped early, the dose level \( d^{**} \) is selected as the MTD which maximizes the posterior probability of toxicity rate falling into the targeted interval i.e., \( d^{**} = \arg\max_{d=1,\ldots,D} \text{Prob}\{p_d \in (p_T - \epsilon_1, p_T + \epsilon_2) \mid \text{Data}\} \) among all doses that are used and do not violate the EWOC rule.
1.3.4 The Modified Toxicity Probability Interval (mTPI) Design

This section describes the modified toxicity probability interval (mTPI) design proposed by Ji et al. (2010). The mTPI design is an extension of the toxicity probability interval (TPI) method (Ji et al., 2007), which uses a simple Bayesian hierarchical model and a decision framework for dose finding.

The mTPI design starts from the specification of three intervals: the under-dosing interval \((0, p_T - \epsilon_1)\), the proper dosing interval \((p_T - \epsilon_1, p_T + \epsilon_2)\) and the over-dosing interval \((p_T + \epsilon_2, 1)\). Unlike the CRM and BLRM, which assumes a parametric curve to model the dose-toxicity response, the mTPI uses a simple beta-binomial model to estimate the toxicity probability and makes the decisions of dose escalation and de-escalation based on the unit probability mass (UPM) of the three intervals. At the end, mTPI selects the dose of which the isotonic transformed toxicity probability is the closest to the target \(p_T\) as the MTD.

1.3.4.1 Probability Model

Consider a phase I trial with \(D\) candidate doses for escalation. Let \(p_1, \ldots, p_D\) denote the true toxicity probabilities for doses \(d = 1, \ldots, D\). The observed data include \(n_d\), the number of patients treated at dose \(d\), and \(y_d\), the number of patients experiencing a toxicity. Let \(Data = \{(y_d, n_d); d = 1, 2, \ldots, D\}\).

The mTPI design employs a simple beta-binomial hierarchical model as follow:

\[
\begin{align*}
    y_d | n_d, p_d &\sim \text{binomial}(n_d, p_d) \\
    p_d &\sim \text{beta}(\alpha, \beta)
\end{align*}
\]

The posterior distribution of \(p_d\) is given by

\[
p_d | y_d, n_d \sim \text{beta}(\alpha + y_d, \beta + n_d - y_d).
\]

In U-Design, we adopt the prior \(\text{beta}(1, 1)\) for \(p_d\), because it would lead to slightly conservative posterior inference as the prior mean is 0.5, which is usually above \(p_T\).

1.3.4.2 Dosing Intervals

The under-dosing interval is defined as \((0, p_T - \epsilon_1)\), the over-dosing interval as \((p_T + \epsilon_2, 1)\), and the equivalence interval as \((p_T - \epsilon_1, p_T + \epsilon_2)\) for proper dosing, where \(\epsilon_1\) and \(\epsilon_2\) are small fractions, such as 0.05, to account for the uncertainty around the true target toxicity \(p_T\). The three dosing intervals are associated with three different dose-finding decisions. The under-dosing interval corresponds
1.3.4 The Modified Toxicity Probability Interval (mTPI) Design

to a dose escalation (E), the over-dosing interval corresponds to a dose de-escalation (D), and the equivalence interval corresponds to staying (S) at the current dose.

1.3.4.3 Dose Finding Rules

Given an interval and a probability distribution, define the UPM of that interval as the probability of the interval divided by the length of the interval. The mTPI design calculates the UPMs for the three dosing intervals, and the one with the largest UPM implies the corresponding dose-finding decision. That decision provides the dose level to be used for future patients. More specifically, given the current dose level \( d \), the mTPI conducts the following steps for dose assignment for the future patients.

1. Compute the UPM for each of the three toxicity probability intervals as follows:

\[
\begin{align*}
\text{UPM(D)}_d &= \frac{\text{Prob}\{p_d \in (p_T + \epsilon_2, 1) \mid \text{Data}\}}{1 - (p_T + \epsilon_2)}, \\
\text{UPM(S)}_d &= \frac{\text{Prob}\{p_d \in (p_T + \epsilon_1, p_T + \epsilon_2) \mid \text{Data}\}}{\epsilon_1 + \epsilon_2}, \\
\text{UPM(E)}_d &= \frac{\text{Prob}\{p_d \in (0, p_T + \epsilon_1) \mid \text{Data}\}}{p_T - \epsilon_1}.
\end{align*}
\]

Here, the numerator in UPM calculation, \( \text{Prob}\{\cdot\} \) is calculated according to the beta posterior distribution in (2.1).

2. Select one of the following actions: “E”, “S” or “D” corresponding to the highest UPM of each toxicity interval. That is, the dose decision is given by

\[
M^* = \arg\max_{M \in \{D, S, E\}} \text{UPM}(M)_d.
\]

In other words,

- Escalate to dose \((d + 1)\), if \( \text{UPM}(E)_d > \text{UPM}(S)_d \) and \( \text{UPM}(E)_d > \text{UPM}(D)_d \),
- Stay at dose \( d \), if \( \text{UPM}(S)_d \geq \text{UPM}(E)_d \) and \( \text{UPM}(S)_d > \text{UPM}(D)_d \),
- De-escalate to dose \((d - 1)\), if \( \text{UPM}(D)_d \geq \text{UPM}(E)_d \) and \( \text{UPM}(D)_d \geq \text{UPM}(S)_d \).

For example, if the under-dosing interval has the largest UPM, decision \( M^* = E \) will be executed and the next cohort of patients will be treated at the next higher dose level \((d + 1)\).

Ji et al. (2010) and Guo et al. (2017) have shown that the above UPM-based decision rules correspond to the Bayes’ rule under a formal Bayesian decision theoretic framework, if we use the uniform prior for \( p_d \).

Your trial designs anywhere, anytime.
Figure 1.24: An example of mTPI decision table generated via U-Design. The target toxicity probability $p_T = 0.3$, and the equivalence interval (EI) is $(0.25, 0.35)$ for up to 18 subjects. Each column represents ($n$) number of subjects treated at the current dose and each row represents ($y$) number of subjects with DLTs at the dose level. Each cell in the table provides the dose-finding decision based on the readouts from the corresponding row ($y$) and column ($n$). The letters in the decision table represent different dose-assignment decisions.

The mTPI design pre-calculates all the dose-finding decisions in advance, allowing investigators to examine the decisions before the trial starts. See Figure 1.24 for an example. Therefore, mTPI exhibits the same simplicity and transparency as rule-based methods like 3+3. The decision table can be generated via U-Design under module Decision & MTD.

1.3.4.4 Design Algorithm

The mTPI algorithm proceeds as follows:

1. At each dose level, treat a cohort of patients, with the first cohort at a prespecified starting dose.

2. After all patients in each cohort complete the DLT evaluation, the dose-finding decision for
1.3. Statistical Methods Review

1.3.4. The Modified Toxicity Probability Interval (mTPI) Design

the next cohort will be determined according to the following rules:

(a) Compute the posterior probability of excessive toxicity at the current tried dose, i.e.,
\[ \text{Prob}\{p_d > p_T | Data\} \]
which is a function of the cumulative distribution of \( \text{beta}(\alpha_0 + y_d, \beta_0 + n_d - y_d) \), similar in (2.1). In U-Design, \( \alpha_0 = \beta_0 = 1 \) is used.

i. **[Additional Safety Rule 1: Dose Exclusion]** If the current dose is considered excessively toxic, i.e., \( \text{Prob}\{p_d > p_T | Data\} > \xi \), where the threshold \( \xi \) is close to 1, say 0.95, the current and all higher doses will be excluded and never be used again in the remainder of the trial to avoid any other patients receiving treatment at those doses.

Also, at that time, the decision is “D”, to de-escalate to previous lower dose.

ii. **[Additional Safety Rule 2: Early Stop]** If the current dose is the lowest dose and considered excessively toxic according to Rule 1 in i, early stop the trial for safety.

(b) If the trial is not stopped early, assign the next cohort of patients to the dose according to the decision table or the procedures in Section 1.3.4.3.

(c) If the dose-assignment decision is “E” but the next higher dose has been excluded by Rule 1, continue to enroll the next cohort at the current dose instead.

(d) If the dose-assignment decision is “E” and the current dose is the highest dose, continue to enroll the next cohort at the current dose instead.

(e) If the dose-assignment is “D” and the current dose is the lowest dose, continue to enroll the next cohort at the current dose instead.

3. Repeat steps 1-2, stop the trial when any of the following conditions is satisfied:

(a) If the prespecified maximum total sample size is reached;

(b) If the lowest dose shows excessive toxicity according to Rule 2; In this case, the trial is early stopped and the MTD cannot be determined;

(c) Optional: ad-hoc rules of maximum number of patients at a dose, denoted by \( K (K < n) \):

- If the mTPI decision is “S”, to stay at the current dose, and the current dose has enrolled \( K \) patients;

- If the mTPI decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled \( K \) patients;

- If the mTPI decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled \( K \) patients.
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

1.3.4.5 MTD Selection

Once all the enrolled patients complete the DLT observation and the trial is not stopped early, an isotonic regression (Ji et al., 2010; Ivanova and Wang, 2006) is used to select the MTD based on the observed DLT data from all the dose levels. Follow the steps below:

1. Compute the isotonically transformed posterior means of DLT probabilities for all the dose levels in the following two steps.

   (a) Using the accumulated safety information about $y_d$ and $n_d$ for $d = 1, \ldots, D$, compute the posterior mean and variance for all the dose levels, $\{\tilde{p}_1, \cdots, \tilde{p}_D\}$ and $\{v_1, \cdots, v_D\}$. Here in U-Design, an independent prior $\text{beta}(0.005, 0.005)$ is used to compute the posterior mean and variance.

   (b) Compute isotonic regression estimates of the posterior mean by solving the optimization problem, minimizing $\sum_{d=1}^{D}(\hat{p}_d - \tilde{p}_d)^2 / v_d$ subject to $\hat{p}_j \geq \hat{p}_k$, for $j > k$. Such optimization can be done using the pooled adjacent violators algorithm (PAVA) (Goodman et al., 1995), the estimated DLT probabilities satisfying the order constraint is obtained, denoted by $\{\hat{p}_1, \cdots, \hat{p}_D\}$.

2. Among all the tried doses for which $\text{Prob}\{p_d > p_T \mid \text{Data}\} < \xi$ and $\hat{p}_d \leq p_T + \epsilon_2$, select as the estimated MTD the dose with the smallest difference $|\hat{p}_d - p_T|$. That is, the estimated MTD is $d^* = \arg\max_d |\hat{p}_d - p_T|$.

3. In case of a tie (i.e., two or more doses have the smallest difference),

   (a) If there is at least one dose lower the target $p_T$ among all the tied doses, choose the highest dose among those as the estimated MTD;

   (b) Otherwise, choose the lowest dose among the tied doses as the estimated MTD.

Your trial designs anywhere, anytime.
1.3. Statistical Methods Review

1.3.5. The Modified Toxicity Probability Interval-2 (mTPI-2) Design

1.3.5.1 The Modified Toxicity Probability Interval-2 (mTPI-2) Design

This section describes the modified toxicity probability interval-2 (mTPI-2) design proposed by Guo et al. (2017). The mTPI-2 improves the mTPI design by blunting the Ockhams razor that leads to some statistically sound but ethical challenging decisions in mTPI. For example, when \( p_T = 0.3 \) and 3 out of 6 patients experience DLTs at a dose, the mTPI decision is “S”, stay at the current dose and enroll more patients. Such a decision may be considered too aggressive. To this end, mTPI-2 constructs a series of dosing interval with equal length to guide the dose escalation and de-escalation, mitigating the effect of interval length in the mTPI design. Otherwise, the model, the design algorithm, and the MTD selection are the same as those in mTPI in Section 1.3.4.

1.3.5.1 The effect of Ockham’s razor in mTPI

The mTPI design has been shown to be simple, transparent, and superior to the 3+3 design (Ji and Wang, 2013; Yang et al., 2015). However, some decisions in mTPI may be debated in practice. For example, when the target toxicity probability \( p_T = 0.3 \), and 3 out of 6 patients treated at a dose experience DLT events, mTPI would suggest “S”, stay at the current dose and enroll more patients to be treated at the dose. Since the empirical rate is \( 3/6 \), or 50%, oftentimes one would argue that the more desirable decision should be D, de-escalate to the next lower dose level. Another case is when \( p_T = 0.3 \) and 2 out of 9 patients experience DLT events at a dose, mTPI would suggest S as well. Investigators could argue that the decision should be E, escalation since the empirical rate is \( 2/9 \), or 22%. Guo et al. (2017) noted that these decisions are due to the Ockham’s razor (Jefferys and Berger, 1992), which is a Bayesian principle that prefers parsimonious models in model selection. The mTPI design treats the three intervals as three models, and penalizes models based on the model size which is the length of each interval. Figure 1.25 gives an example of the effect of the Ockham’s razor in mTPI. Statistically speaking, there is nothing wrong with the Ockham’s razor in mTPI as the Bayesian inference takes into account the model complexity when choosing the optimal decision. However, for human clinical trials patient safety often outweighs statistical optimality. To this end, mTPI-2 modifies the decision theoretic framework and blunt the Ockham’s razor, which leads to practically desirable decision rules.
Figure 1.25: An example demonstrating the effect of the Ockham’s razor in mTPI. Shown is the posterior density of $p_d$ when $x_d = 3$ and $n_d = 6$. Even though the shape of the density suggests that dose $d$ might be above the MTD, e.g., the posterior mode is to the right of the equivalence interval (shown as the two vertical bars), the UPM for decision “S” (stay) is still larger than that of the UPM for decision D (de-escalate). Therefore, mTPI would still choose to “Stay” despite that the shape of the posterior density of $p_d$ indicates otherwise. This is due to the larger size (longer length) of the interval $M_D$ than $M_S$ and the Ockham’s razor, which prefers the smaller model $M_S$. 
1.3. Statistical Methods Review
1.3.5. The Modified Toxicity Probability Interval-2 (mTPI-2) Design

1.3.5.2 Dose Finding Rules

The basic idea in mTPI-2 is to divide the unit interval \((0, 1)\) into subintervals with equal length, given by \((\epsilon_1 + \epsilon_2)\). This results in multiple intervals with the same length except for the boundary intervals, see Figure 1.26. For clarify, denote \(EI\) the equivalence interval \(\left(p_T - \epsilon_1, p_T + \epsilon_2\right)\), and \(LI\) a set of intervals below \(EI\), and \(HI\) a set of intervals above \(EI\). For example, when \(p_T = 0.3\) and \(\epsilon_1 = \epsilon_2 = 0.05\), the \(EI = (0.25, 0.35)\), the \(LI\) intervals are

\[
LI = \{M_{1L}^{LI} = (0.15, 0.25), M_{2L}^{LI} = (0.05, 0.15), M_{3L}^{LI} = (0, 0.05)\},
\]

and the \(HI\) intervals are

\[
HI = \{M_{1H}^{HI} = (0.35, 0.45), M_{2H}^{HI} = (0.45, 0.55), M_{3H}^{HI} = (0.55, 0.65), M_{4H}^{HI} = (0.65, 0.75), M_{5H}^{HI} = (0.75, 0.85), M_{6H}^{HI} = (0.85, 0.95), M_{7H}^{HI} = (0.95, 1)\}.
\]

Other than the boundaries \((0, 0.05)\) and \((0.95, 1)\), all the intervals have the same length. The boundaries do not affect the decision making since they are clearly associated with “E” and “D” decisions, respectively. See Guo et al. (2017) for details.

The dose finding rules are given as follows:

- If the equivalence interval \(M^{EI} = (p_T - \epsilon_1, p_T + \epsilon_2)\) has the largest UPM, it is selected as the winning interval and dose-assignment decision of mTPI-2 is “S”, to stay at the current dose.
- If any interval \(M_{j}^{LI}\) in \(LI\) has the largest UPM, it is selected as the winning interval and dose-assignment decision of mTPI-2 is “E”, to escalate to the next higher dose.
- If any interval \(M_{k}^{HI}\) in \(HI\) has the largest UPM, it is selected as the winning interval and dose-assignment decision of mTPI-2 is “D”, to de-escalate to the previous lower dose.

In Figure 1.26, for the same posterior density corresponding to \(y_d = 3\) and \(n_d = 6\), interval \(M_{2}^{HI}\) exhibits the largest UPM and therefore the decision is now “D”. Note that the same decision theoretic framework as mTPI is in place except that now there are multiple intervals corresponding to “D” or “E”, and the intervals all have the same length except the boundary ones, thereby blunting the Ockham’s razor.

The same as mTPI, all the dose-finding decisions of mTPI-2 can be pre-tabulated in advance, allowing investigators to examine the decisions before the trial starts. see Figure 1.27 for an example. And the decision table can also be generated via U-Design under module Decision & MTD.
Figure 1.26: An example demonstrating the new framework of mTPI-2. Here, \( EI \) is the equivalence interval \( (p_T - \epsilon_1, p_T + \epsilon_2) \), and \( LI \) denotes the intervals below \( EI \), and \( HI \) denotes the intervals above \( EI \). Interval \( M_{2}^{HI} \) exhibits the largest UPM and therefore the decision is now “D”, to de-escalate.
1.3. Statistical Methods Review

1.3.5. The Modified Toxicity Probability Interval-2 (mTPI-2) Design

**Figure 1.27:** An example of mTPI-2 decision table generated via U-Design. The target toxicity probability \( p_T = 0.3 \), and the equivalence interval (EI) is \((0.25, 0.35)\) for up to 18 subjects. Each column represents \((n)\) number of subjects treated at the current dose and each row represents \((y)\) number of subjects with DLTs at the dose level. Each cell in the table provides the dose-finding decision based on the readouts from the corresponding row \((y)\) and column \((n)\). The letters in the decision table represent different dose-assignment decisions.

### 1.3.5.3 The Keyboard Design

The Keyboard design is proposed by Yan et al. (2017), which is based on the same construction as the mTPI-2. In the Keyboard design, the sub-intervals are called “keys” and the key associated with the largest posterior probability is chosen to guide the dose-assignment decisions. When the intervals are with equal-length, the winning interval with the largest posterior probability is the same as the interval with the largest UPM. Therefore, the keyboard design is the same as the mTPI-2 design.
1.3.6 The Modified Cumulative Cohort Design (mCCD)

The cumulative cohort design (CCD) was formally proposed by Ivanova et al. (2007), which is also an interval-based design. But unlike the mTPI and mTPI-2 designs (in Sections 1.3.4 and 1.3.5, respectively) which calculate the posterior probability that the toxicity rate \( p_d \) falls into each interval and decide the decision based on a formal Bayesian decision framework, the CCD design just relies on the point estimate \( \hat{p}_d \) and compares it with the equivalence interval boundaries, \((p_T - \epsilon_1)\) and \((p_T + \epsilon_2)\). In U-Design, we construct a modified CCD (mCCD) design, which follows the same concept for dose finding as CCD, except that we add some other safety rules. The mCCD design is not published and is a property of Laiya Consulting, Inc.

1.3.6.1 Probability Model

Consider a phase I trial with \( D \) candidate doses for escalation. Let \( p_1, \ldots, p_D \) denote the true toxicity probabilities for doses \( d = 1, \ldots, D \). The observed data include \( n_d \), the number of patients treated at dose \( d \), and \( y_d \), the number of patients experiencing a toxicity. Let \( Data = \{(y_d, n_d); d = 1, 2, \ldots, D\} \).

The CCD design simply uses the empirical point estimate \( \hat{p}_d = y_d/n_d \) as the estimation of toxicity rate \( p_d \) for dose level \( d \).

1.3.6.2 Dosing Intervals

The mCCD design prespecifies three toxicity probability intervals: the under-dosing interval \((0, p_T - \epsilon_1)\), the equivalence interval \((p_T - \epsilon_1, p_T + \epsilon_2)\), and the over-dosing interval \([p_T + \epsilon_2, 1)\), where \( \epsilon_1 \) and \( \epsilon_2 \) are small fractions, such as 0.05, to account for the uncertainty around the true target toxicity. The three dosing intervals are associated with three different dose-escalation decisions. The under-dosing interval corresponds to a dose escalation (E), the over-dosing interval corresponds to a dose de-escalation (D), and the equivalence interval corresponds to staying at the current dose (S).
1.3.6. The Modified Cumulative Cohort Design (mCCD)

Figure 1.28: An example of mCCD decision table generated via U-Design. The target toxicity probability $p_T = 0.3$, and the equivalence interval (EI) is $(0.25, 0.35)$ for up to 18 subjects. Each column represents ($n$) number of subjects treated at the current dose and each row represents ($y$) number of subjects with DLTs at the dose level. Each cell in the table provides the dose-finding decision based on the readouts from the corresponding row ($y$) and column ($n$). The letters in the decision table represent different dose-assignment decisions.

### 1.3.6.3 Dose Finding Rules

Suppose the current dose level is $d$, the mCCD applies the same concept for dose finding as CCD, that is, uses the equivalence interval as the boundaries for thresholding the estimate $\hat{p}_d$. Specifically,

1. **Escalate to dose** $(d + 1)$, if $\hat{p}_d \in (0, p_T - \epsilon_1]$, i.e., if $y_d/n_d \leq p_T - \epsilon_1$,
2. **Stay at dose** $d$, if $\hat{p}_d \in (p_T - \epsilon_1, p_T + \epsilon_2)$, i.e., if $p_T - \epsilon_1 < y_d/n_d < p_T + \epsilon_2$,
3. **De-escalate to dose** $(d - 1)$, if $\hat{p}_d \in (p_T + \epsilon_2, 1]$, i.e., if $y_d/n_d \geq p_T + \epsilon_2$.

The decision table based on the above rules can be generated via U-Design before the beginning of the trial for investigators to examine. see Figure 1.28 for an example.
1.3.6.4 Design Algorithm

The mCCD algorithm is similar as mTPI, which proceeds as follows:

1. At each dose level, treat a cohort of patients, with the first cohort at a prespecified starting dose.
2. After all patients in each cohort complete the DLT evaluation, the dose-assignment decision for the next cohort will be determined according to the following rules:
   
   (a) Compute the posterior probability of excessive toxicity at the current tried dose, i.e.,
   \[ \text{Prob}\{ p_d > p_T | \text{Data} \} \]
   which is a function of the cumulative Beta distribution \( \text{Beta}(\alpha_0 + y_d, \beta_0 + n_d - y_d) \). In U-Design, \( \alpha_0 = \beta_0 = 1 \) is used.

   i. [Additional Safety Rule 1: Dose Exclusion] If the current dose is considered excessively toxic, i.e., \( \text{Prob}\{ p_d > p_T | \text{Data} \} > \xi \), where the threshold \( \xi \) is close to 1, say 0.95, the current and all higher doses are excluded and never be used again in the remainder of the trial.
   Also, at that time, the decision is “D”, to de-escalate to previous lower dose.

   ii. [Additional Safety Rule 2: Early Stop] If the current dose is the lowest dose and considered excessively toxic according to Rule 1 in i, early stop the trial for safety.

   (b) If the trial is not stopped early, assign the next cohort of patients to the dose according to the decision table or the procedures in Section 1.3.6.3.

   (c) If the dose-assignment decision is “E” but the next higher dose has been excluded by Rule 1, continue to enroll the next cohort at the current dose instead.

   (d) If the dose-assignment decision is “E” and the current dose is the highest dose, continue to enroll the next cohort at the current dose instead.

   (e) If the dose-assignment is “D” and the current dose is the lowest dose, continue to enroll the next cohort at the current dose instead.

3. Repeat steps 1-2, stop the trial when any of the following conditions is satisfied:

   (a) If the prespecified maximum total sample size is reached;

   (b) If the lowest dose shows excessive toxicity according to Rule 2; In this case, the trial is early stopped and the MTD cannot be determined;

   (c) Optional: ad-hoc rules of maximum number of patients at a dose, denoted by \( K (K < n) \):

   - If the mCCD decision is “S”, to stay at the current dose, and the current dose has enrolled \( K \) patients;
1.3.6. The Modified Cumulative Cohort Design (mCCD)

- If the mCCD decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled \( K \) patients;
- If the mCCD decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled \( K \) patients.

1.3.6.5 MTD Selection

Once all the enrolled patients complete the DLT observation and the trial is not stopped early, an isotonic regression (Ji et al., 2010; Ivanova and Wang, 2006) is used to select the MTD based on the observed DLT data from all the dose levels. Follow the steps below:

1. Compute the isotonically transformed posterior means of DLT probabilities for all the dose levels in the following two steps.

   (a) Using the accumulated safety information about \( y_d \) and \( n_d \) for \( d = 1, \ldots, D \), compute the posterior mean and variance for all the dose levels, \( \{\tilde{p}_1, \ldots, \tilde{p}_D\} \) and \( \{v_1, \ldots, v_D\} \). Here in U-Design, an independent prior \( \text{Beta}(0.005, 0.005) \) is used to compute the posterior mean and variance.

   (b) Compute isotonic regression estimates of the posterior mean by solving the optimization problem, minimizing \( \sum_{d=1}^{D} (\hat{p}_d - \tilde{p}_d)^2/v_d \) subject to \( \hat{p}_j \geq \hat{p}_k \), for \( j > k \). Such optimization can be done using the pooled adjacent violators algorithm (PAVA) (Goodman et al., 1995), the estimated DLT probabilities satisfying the order constraint is obtained, denoted by \( \{\hat{p}_1, \ldots, \hat{p}_D\} \).

2. Among all the tried doses for which \( \text{Prob}\{p_d > p_T \mid Data\} < \xi \), select as the estimated MTD the dose with the smallest difference \( |\hat{p}_d - p_T| \). That is, the estimated MTD is \( d^* = \arg\max_d |\hat{p}_d - p_T| \).

3. In case of a tie (i.e., two or more doses have the smallest difference),
   
   (a) If there is at least one dose lower the target \( p_T \) among all the tied doses, choose the highest dose among those as the estimated MTD;
   
   (b) Otherwise, choose the lowest dose among the tied doses as the estimated MTD.
1.3.6.6 Differences with the BOIN Design

Liu and Yuan (2015a) extended CCD and developed the BOIN design, with local and global BOIN as two versions. They stated that BOIN is an improvement of CCD since it uses interval boundaries that are optimal based on an optimal criterion. In BOIN, the dose-finding problem is cast as an optimization problem. Denote as $\phi_1 = p_T - a$ and $\phi_2 = p_T + b$ two user-provided values which represent the lower and upper bound of the equivalence interval for $p_T$. In this sense, $\phi_1$ and $\phi_2$ are given and decided by the trial investigators, which is similar to $(p_T - \epsilon_1)$ and $(p_T + \epsilon_2)$ in the mTPI and mTPI-2 designs. Suppose dose $d$ is currently administered in the trial. In the local BION design, an ad-hoc optimization procedure is defined to minimize the probability of making an erroneous decision, treating $\phi_1$ and $\phi_2$ as two parameters. Then the optimal values are obtained and denoted as $\lambda_1$ and $\lambda_2$, given by

$$
\lambda_1 = \log \left( \frac{1 - \phi_1}{1 - p_T} \right) / \left( \frac{p_T(1 - \phi_1)}{\phi_1(1 - p_T)} \right)
$$

$$
\lambda_2 = \log \left( \frac{1 - p_T}{1 - \phi_2} \right) / \left( \frac{\phi_2(1 - p_T)}{p_T(1 - \phi_2)} \right)
$$

The BOIN design first examines if $\hat{p}_d$ falls into one of the three intervals $(0, \lambda_1]$, $(\lambda_1, \lambda_2]$, $[\lambda_2, 1)$, and escalates to dose $(d + 1)$, stays at dose $d$, or de-escalates to dose $(d - 1)$, accordingly. In other words, the BOIN design uses the same concept for dose finding as the CCD design, except BOIN changes the original user-provided boundary $\phi_{1,2}$ to $\lambda_{1,2}$ based on an optimization criterion. And Liu and Yuan (2015a) showed that $(\lambda_1, \lambda_2)$ is always nested in the original interval $(\phi_1, \phi_2)$ under local BOIN. In contrast, the mTPI (mTPI-2) and mCCD designs do not have the $\lambda$s and use the user-provided $\phi$’s values for decision making. See Figure 1.29 for an illustration (Ji and Yang, 2017). As the Figure 1.29 shown, there is a gap between the $\lambda$s and the user-provided $\phi$’s values in BOIN’s decision making. The gap is small enough to be ignorable when the sample size is small. So BOIN performs well in terms of operating characteristics of safety and reliability in phase I trials.

In U-Design, we currently do not implement BOIN. Instead, we implement the modified version of CCD, which uses the user-provided $\phi$’s (i.e., $\epsilon$’s) for decision making. But the mCCD is essentially the same as BOIN_{lambda} described in Ji and Yang (2017), which has been shown to exhibit desirable operating characteristics. We do not wish to change the user-provided values on $\phi$’s and do not believe that they can be optimized.
The mTPI (mTPI-2) Design

The BOIN Design

Figure 1.29: A graphical illustration between the decision frameworks under the mTPI (mTPI-2) and the BOIN design. Under mTPI, the two probability boundaries $\phi_1 = p_T - a$ and $\phi_2 = p_T + b$ are elicited from the clinicians and treated as known. Under BOIN, the two $\phi_1$ and $\phi_2$ values are also elicited, but not used for decision making. Instead, two new values, $\lambda_1$ and $\lambda_2$ are derived based on an optimization procedure and used for decision making. There is a gap on each side of the $p_T$ (right panel) due to the optimization process, and the gap is independent of sample size.
1.3.7 The i3+3 Design

The i3+3 design is a rule-based design for finding the maximum tolerated dose (MTD) proposed by Liu et al. (2020). The i3+3 design defines an equivalence interval (EI) \([p_T - \epsilon_1, p_T + \epsilon_2]\) with the target probability of toxicity \(p_T\) and two small fractions, \(\epsilon_1\) and \(\epsilon_2\), and allocates the next cohort of patients based on the relationship between toxicity rate observed on the current cohort of patients and the equivalence interval. Similar to the 3+3 design, i3+3 is rule-based but assumes that toxicity increases with dose. It has been demonstrated to perform as good as major model-based designs and is flexible enough to accommodate different target toxicity probability as well as different cohort sizes (Liu et al., 2020).

1.3.7.1 Design Algorithm

**Dose finding rules:** Suppose dose \(d\) is currently used in the trial to treat patients, and \(y_d\) patients have experienced dose limiting toxicities (DLTs) out of \(n_d\) patients that have been treated. Based on EI, the i3+3 design identifies the appropriate dose for the next cohort of patients according to the following five simple rules, which accounts for the variability in the observed toxicity data (\(y_d\) and \(n_d\)) for each dose.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Decision</th>
<th>Next dose level</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\frac{y_d}{n_d}) below EI</td>
<td>Escalation(E)</td>
<td>(d + 1)</td>
</tr>
<tr>
<td>(\frac{y_d}{n_d}) inside EI</td>
<td>Stay(S)</td>
<td>(d)</td>
</tr>
<tr>
<td>(\frac{y_d}{n_d}) above EI and (\frac{y_d - 1}{n_d}) below EI</td>
<td>Stay(S)</td>
<td>(d)</td>
</tr>
<tr>
<td>(\frac{y_d}{n_d}) above EI and (\frac{y_d - 1}{n_d}) inside EI</td>
<td>De-escalation(D)</td>
<td>(d - 1)</td>
</tr>
<tr>
<td>(\frac{y_d}{n_d}) above EI and (\frac{y_d - 1}{n_d}) above EI</td>
<td>De-escalation(D)</td>
<td>(d - 1)</td>
</tr>
</tbody>
</table>

Here, a value is below the EI means that the value is smaller than \((p_T - \epsilon_1)\), the lower bound of the EI. A value is inside the EI means that the value is larger than or equal to \((p_T - \epsilon_1)\) but smaller than or equal to \((p_T + \epsilon_2)\). A value is above the EI mean that the value is larger than \((p_T + \epsilon_2)\), the upper bound of the EI. All potential decisions based on the above set of rules could be pre-tabulated in advance via U-Design under module **Decision & MTD**, allowing investigators for examination before the trial starts. See Figure 1.30 for an illustration. When \(d\) is the highest dose or lowest dose, the above rules are modified as special cases:

- If the current dose is the highest dose, and \(\frac{y_d}{n_d}\) is below the EI, stay (“S”) instead of escalating (“E”) because there is no dose to escalate to.
1.3. Statistical Methods Review

1.3.7. The i3+3 Design

Figure 1.30: An example of i3+3 decision table generated via U-Design. The target toxicity probability $p_T = 0.3$, and the equivalence interval (EI) is $(0.25, 0.35)$ for 18 subjects. Each column represents ($n$) number of subjects treated at the current dose and each row represents ($y$) number of subjects with DLTs at the dose level. Each cell in the table provides the dose-finding decision based on the readouts from the corresponding row ($y$) and column ($n$). The letters in the decision table represent different dose-assignment decisions.

- If the current dose is the lowest dose, and $\frac{y}{n}$ is above the EI, stay (“S”) instead of potentially de-escalating (“D”) because there is no dose to de-escalate to.

Safety rules: Following the mTPI and mTPI-2 design (Ji et al., 2010; Ji and Wang, 2013; Guo et al., 2017), two safety rules are added as ethical constraints to avoid excessive toxicity:

- [Rule 1: Dose Exclusion] If the current dose is considered excessively toxic, i.e., $\text{Prob}\{p_d > p_T | \text{Data}\} > \xi$, where the threshold $\xi$ is close to 1, say 0.95, the current and all higher doses are excluded and never be used again in the remainder of the trial.

- [Rule 2: Early Stop] If the current dose is the lowest dose (first dose) and is considered excessively toxic according to Rule 1, early stop the trial for safety.

In safety Rules 1 and 2, $\text{Prob}\{p_d > p_T | \text{Data}\}$ is a function of the cumulative beta distribution
Module 1. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

\[ beta(\alpha_0 + y_d, \beta_0 + n_d - y_d), \]  and \( \alpha_0 = \beta_0 = 1 \) is used in U-Design by default. And if i3+3 decision based on the current dose is “E”, i.e., \( \frac{y_d}{n_d} \) is below the EI, while the next higher dose level \((d+1)\) has been declared excessive toxicity and been excluded, stay (“S”) instead of escalating (“E”) because there is no available dose to escalate to.

**Trial termination:** The trial proceeds until any of the following stopping criteria is met:

1. If the prespecified maximum total sample size \( n \) is reached;
2. If the lowest dose shows excessive toxicity according to Rule 2; In this case, the trial is early stopped and the MTD cannot be determined;
3. Optional: ad-hoc rules of maximum number of patients in one dose, denoted by \( K(K < n) \):
   - If the i3+3 decision is “S”, to stay at the current dose, and the current dose has enrolled \( K \) patients;
   - If the i3+3 decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled \( K \) patients;
   - If the i3+3 decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled \( K \) patients;

### 1.3.7.2 MTD Selection

Once all the enrolled patients complete the DLT observation and the trial is not stopped early, the MTD selection under the i3+3 design follows the same procedure as in the mTPI and mTPI-2 design (Ji et al., 2010; Ji and Wang, 2013; Guo et al., 2017). Follow the steps below:

1. Compute the isotonically transformed posterior means of DLT probabilities for all the dose levels in the following two steps.

   (a) Using the accumulated safety information about \( y_d \) and \( n_d \) for \( d = 1, \ldots, D \), compute the posterior mean and variance for all the dose levels, \( \{\hat{p}_1, \ldots, \hat{p}_D\} \) and \( \{v_1, \ldots, v_D\} \). Here in U-Design, an independent prior \( beta(0.005, 0.005) \) is used to compute the posterior mean and variance.

   (b) Compute isotonic regression estimates of the posterior mean by solving the optimization problem, minimizing \( \sum_{d=1}^{D}(\hat{p}_d - \tilde{p}_d)^2 / v_d \) subject to \( \hat{p}_j \geq \hat{p}_k \), for \( j > k \). Such optimization can be done using the pooled adjacent violators algorithm (PAVA) (Goodman et al., 1995), the estimated DLT probabilities satisfying the order constraint is obtained, denoted by \( \{\hat{p}_1, \ldots, \hat{p}_D\} \).
2. Among all the tried doses for which \( \text{Prob}\{p_d > p_T \mid Data\} < \xi \) and \( \hat{p}_d \leq p_T + \epsilon_2 \), select as the estimated MTD the dose with the smallest difference \( |\hat{p}_d - p_T| \). That is, the estimated MTD is \( d^* = \text{argmax}_d |\hat{p}_d - p_T| \).

3. In case of a tie (i.e., two or more doses have the smallest difference),
   
   (a) If there is at least one dose lower the target \( p_T \) among all the tied doses, choose the highest dose among those as the estimated MTD;
   
   (b) Otherwise, choose the lowest dose among the tied doses as the estimated MTD.
2. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment

2.1 Introduction

Phase I oncology dose-finding trials assign cancer patients to ascending doses of a new investiga-
tional drug (or drug combinations) and adaptively decide the dose level of newly enrolled patients
based on observed binary dose-limiting toxicity (DLT) outcomes. The goal is to determine the
maximum tolerated dose (MTD) of the drug(s), defined as the highest dose that has a toxicity prob-
ability less than or close to a prespecified target rate $p_T$. Most popular statistical designs, such as
the 3+3 (Storer, 1989), CRM (O’Quigley et al., 1990), mTPI-2 (Guo et al., 2017), and i3+3 (Liu
et al., 2020) designs described in the Single-Agent Dose-Finding Designs with Toxicity Endpoint
and Rolling Enrollment module, typically enroll patients in cohorts and apply sequential decisions
that determine the dose level for each cohort based on observed toxicity data. Accrual is suspended
after enrollment of each cohort of patients until all the patients in the current cohort have observed
outcomes, with or without DLTs. This type of cohort-based designs can be inefficient, especially if
the trial needs to be frequently suspended. See Skolnik et al. (2008) and Doussau et al. (2016) for
discussion. For example, subsequent patients can be turned away during trial suspension, resulting
in waste of precious patient resource. In addition, trial duration is prolonged due to between-cohort
suspension.

To shorten the study conduct timeline of phase I trials and reduce the number of accrual suspen-
sions, this module describe some rolling-based designs, which allows concurrent patient enrollment
that is faster than cohort-base enrollment, including the rolling six (Skolnik et al., 2008) and R-TPI
designs (Guo et al., 2019). Besides, mTPI-2 (Guo et al., 2017) and 3+3 (Storer, 1989) designs, with
the ethics constraint of “decision-in-advance” applicable to the real-life trials, are also included for
comparison. See more technical details in Section 2.3.
Besides the operating characteristics in terms of the safety and reliability reported in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module, this module enables users to compare the trial duration based on real-life settings, which are characterized as three user-input parameters, the mean inter-patient arrival time, the maximum DLT follow-up period, and the probability of inevaluability (such as drop off) of enrollment patients. The procedure of simulating patients enrollment and evaluation is described in details in Section 2.3.1.

Hereinafter, the terms “Enrollment” and “Accrual” are used interchangeably.
2.2 User Interface and Tutorial

2.2.1 Overview

Entering the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment page, users will see four main tabs: Simulation Setup, Simulation Results, Decision Table and MTD Estimation. The first two tabs allow users to conduct simulations and visualize/download simulation results, and the next two tabs allow users to generate decision tables and estimate the MTD, respectively. In the Simulation Setup tab, there are four steps (Figure 2.1): 1) Set enrollment parameters, 2) Set trial parameters, 3) Select designs, and 4) Generate scenarios. Users need to complete the steps 1-4 to set up simulations for a single design or multiple designs. Upon completing steps 1-4, users click the “Launch Simulation” button at the bottom of the page. Users may also click the “Reset” button next to Launch Simulation to clear all settings. After the simulation is launched, the results of simulations will be displayed in the Simulation Results tab. The simulation process can be monitored in real time at the top of the Simulation Results tab. Detailed steps of using this module are elaborated next in §2.2.2-§2.2.5.
2.2. User Interface and Tutorial

2.2.1. Overview

Figure 2.1: Simulation setup in the **Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment**

Your trial designs anywhere, anytime.
Module 2. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment

2.2.2 Simulation Setup

In the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module, U-Design provides four designs, mTPI-2, 3+3, Rolling 6 and R-TPI, for simulation. Users can choose up to design configurations for simultaneous comparison in the Simulation Setup tab each time. A design configuration means a design such as R-TPI, along with the designs settings, such as sample size. Request to allow more than four design configurations by emailing admin@laiyaconsulting.com.

2.2.2.1 Step 1: Set enrollment parameters

Specify the maximum follow-up time ($T_{\text{follow-up}}$), mean interpatient arrival time ($MIAT$) and inevaluable rate ($IR$) for the enrollment simulation. See Figure 2.2. Hover mouse over each trial parameter, and a description will be displayed explaining the meaning of the parameter. The detailed explanation of the above three input arguments is provided in Table 2.1. The technical details of simulating patients enrollment is provided in §2.3.1.

![Step 1: Set enrollment parameters](image)

**Figure 2.2:** Set enrollment parameters in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment.

**Table 2.1:** Input arguments for enrollment parameters in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{follow-up}}$</td>
<td>The maximum follow-up time</td>
<td>The DLT observation period for each patient in the trial (days). Default value is 21 days.</td>
</tr>
<tr>
<td>$MIAT$</td>
<td>Mean interpatient arrival time</td>
<td>The mean chronologic time (days) for a patient to arrive in the clinic and be eligible for study. Default value is 10 days.</td>
</tr>
<tr>
<td>$IR$</td>
<td>Inevaluable rate</td>
<td>The proportion of patients who entered the trial and received the treatment, but dropped out due to non-DLT related event when being followed up. Default value is 0.1.</td>
</tr>
</tbody>
</table>
2.2. User Interface and Tutorial

2.2.2. Simulation Setup

2.2.2.2 Step 2: Set trial parameters

Specify the target toxic probability \((p_T)\), number of simulations \((n_{\text{sim}})\), and random seed of simulation \((R_{\text{seed}})\) for the simulated trials. See Figure 2.3. Hover mouse over each trial parameter, and a description will be displayed explaining the meaning of the parameter. The detailed explanation of the above three input arguments is provided in Table 2.2.

![Step 2: Set trial parameters](attachment:image)

**Figure 2.3:** Set trial parameters in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_T)</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). The main objective of phase I clinical trials is to find the highest dose with a toxicity probability closest to or lower than (p_T). Default value is 0.3.</td>
</tr>
<tr>
<td>(n_{\text{sim}})</td>
<td>The number of simulated trials</td>
<td>The maximum number of simulated trials allowed is 10,000. Default value is 1,000.</td>
</tr>
<tr>
<td>(R_{\text{seed}})</td>
<td>The random seed of simulation</td>
<td>A random seed is a number used to initialize a pseudorandom number generator in the simulation. Default value is 32432.</td>
</tr>
</tbody>
</table>

Table 2.2: Input arguments for trials in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment.
2.2.2.3 Step 3: Select designs

To select a design, click the button with the design’s name on it. Up to four design configurations may be selected for comparison.

When setting the sample size $n$ for the mTPI-2 and R-TPI designs, two options are provided: 1) match with 3+3, if a 3+3 design is selected; 2) manually input. Check the “Match with 3+3” box to use the average sample size of the selected 3+3 design as the maximum sample size $n$. If two or more 3+3 design configurations are selected, U-Design would choose the first 3+3 design in the design list as the benchmark. Figure 2.4 is an example if one selects mTPI-2, 3+3, rolling six and R-TPI designs, with the sample size of mTPI-2 matching with 3+3 and the sample size of R-TPI being a manually input value, 30.

Click the “Delete” button to remove the selected designs.

Design parameters can be modified in the input box of corresponding row. Hover mouse over each design parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 2.3.

![Step 3: Select designs](image)

Figure 2.4: Select designs in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment.
Table 2.3: Input parameters for designs in the **Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment**.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{\text{start}}$ (all designs)</td>
<td>Starting dose level</td>
<td>The starting dose level in the simulated trials. Default value is 1.</td>
</tr>
<tr>
<td>$n$ (all designs)</td>
<td>Sample size</td>
<td>The maximum number of patients to be treated in the trial. The upper limit is set at 100 since the number of patients that are enrolled in phase I clinical trial is typically small. Default value is 30.</td>
</tr>
<tr>
<td>$\epsilon_1, \epsilon_2$ (mTPI-2, R-TPI)</td>
<td>$\epsilon_1$: lower margin $\epsilon_2$: higher margin</td>
<td>Two small fractions used to define the equivalence/target interval of the MTD. Any dose with a toxicity probability falling into the interval $[p_T - \epsilon_1, p_T + \epsilon_2]$ is considered an acceptable dose of MTD. Default values for both are 0.05.</td>
</tr>
<tr>
<td>$n_{\text{cohort}}$ (mTPI-2)</td>
<td>Cohort size</td>
<td>The number of patients in each cohort. Default value is 3. For 3+3, the cohort size is 3 by default, and for Rolling 6 and R-TPI designs, there is no concept of cohort size and patients are enrolled as needed without suspension.</td>
</tr>
<tr>
<td>$C$ (R-TPI)</td>
<td>The maximum number of pending patients allowed in the trial</td>
<td>The maximum number of pending patients allowed in the trial. It can be provided by users to control the enrollment speed. For rolling six design, $C$ is 6 by default.</td>
</tr>
</tbody>
</table>
2.2.2.4 Step 4: Generate scenarios

There are two ways to generate scenarios, automatically (in below Auto Generation tab, see Figure 2.5) or through manual construction (in below Manual Construction tab, see Figure 2.6). Once scenarios are generated, click the “Launch Simulation” button at the bottom of the page to run \( n_{\text{sim}} \) (set in step 2) simulations, for each scenario and selected design (set in step 3) combination, assuming \( p_T \) (set in step 2) and patients enrollment and follow-up conditions, \( T_{\text{follow-up}} \), MIAT and \( IR \) (set in step 1).

**Auto Generation (Figure 2.5)**

Select the number of doses \( n_{\text{dose}} \) (\( 3 \leq n_{\text{dose}} \leq 10 \)) from the dropdown box. Upon selection, five or six scenarios will be created automatically, each of which contains the true toxicity probabilities for \( n_{\text{dose}} \) dose levels. These generated scenarios are displayed and editable. The detailed algorithm for scenarios auto generation is provided next.

**Manual Construction (Figure 2.6)**

Follow the instructions below to manually construct scenarios. Then click the “Add” button to create these scenarios. The format of input must comply with the following instructions.

- Scenarios should be separated by line breaks;
- The parameters in one scenario should be ordered in accordance with this sequence:
  
  Target toxicity probability, Number of simulated trials, True toxicity probabilities of all the dose levels;

- Each parameter must be separated by a white space or comma.

For example, by inputting “0.2 1000 0.05 0.1 0.15 0.2” or “0.2,1000,0.05,0.1,0.15,0.2”, a scenario is presented that runs 1000 simulated trials with a target \( p_T = 0.2 \) and true toxicity probabilities of four dose levels, 0.05, 0.1, 0.15 and 0.2.

The generated scenarios are displayed as a list (Figures 2.5 and 2.6) which appears below the generation section. An interactive chart will also be generated to visually display the shape of true toxicity probabilities for each scenario.
2.2. User Interface and Tutorial

2.2.2. Simulation Setup

**Figure 2.5:** Automatically generate scenarios in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

**Figure 2.6:** Manually generate scenarios in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.
Algorithm for Auto Generation

By entering the number of candidate dose levels $n_{\text{dose}}$, five or six scenarios are generated automatically. See Figure 2.7 for an illustration. They represent the four types of dose-response shapes below.

<table>
<thead>
<tr>
<th>Types</th>
<th>Dose-Response Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal</td>
<td>Some doses are tolerable but some are overly toxic, AND there exists at least one dose level close to the target $p_T$ or falling within the equivalence interval $[p_T - \epsilon_1, p_T + \epsilon_2]$.</td>
</tr>
<tr>
<td>Safe</td>
<td>All doses are safe and tolerable with the true toxicity probabilities smaller than the target $p_T$ or the lower boundary of equivalence interval $(p_T - \epsilon_1)$.</td>
</tr>
<tr>
<td>Toxic</td>
<td>All doses are overly toxic with the true toxicity probabilities larger than the target $p_T$ or the upper boundary of equivalence interval $(p_T + \epsilon_2)$.</td>
</tr>
<tr>
<td>Steep</td>
<td>Some doses are tolerable but some are overly toxic, AND there is a steep jump in the toxicity probability between two adjacent doses, AND there is no dose close to the target $p_T$ or falling within the equivalence interval $[p_T - \epsilon_1, p_T + \epsilon_2]$.</td>
</tr>
</tbody>
</table>

Two “Steep” scenarios are generated, with the toxicity probability steep jump occurring at the first or second half of the doses.

Similarly, two “Ideal” scenarios might be generated, with the MTD placed in the first or second half of the doses.
2.2. User Interface and Tutorial

2.2.2. Simulation Setup

![Figure 2.7: An example of four main types of scenarios. Five dose levels are considered for escalation. The target toxicity probability is $p_T = 0.25$, and the equivalence interval is $EI=[0.2, 0.3]$. Six different lines represent four main types of scenario, respectively. In “Ideal” scenarios (Lines 1 and 2), doses 2 and 4 are the true MTD with toxicity probability falling within the EI, respectively. In “Safe” scenario (Line 3), all doses are safe with toxicity probabilities smaller than the target $p_T = 0.25$. “Toxic” scenario (Line 4) gives a contrary situation to the “Safe” scenario, where all doses are overly toxic with the toxicity probabilities larger than the target $p_T = 0.25$. The remaining two lines (Lines 5 and 6) are “Steep” scenario, in which some doses are tolerable but some are overly toxic and there is a steep jump in the toxicity probability occurring at the first or second half of the doses (between doses 4 and 5 in Line 5, and between doses 1 and 2 in Line 6).]
Module 2. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment

2.2.2.5 Launch simulation

Once the steps 1-4 are completed, users can conduct simulated clinical trials to examine the operating characteristics of the selected designs using the selected scenarios, by clicking the “Launch Simulation” button at the bottom of Simulation Setup tab (Figures 2.5 and 2.6). A green “Launch Successful” message will be displayed on the website as in Figure 2.8 to indicate that the simulation has been successfully launched. Users may click the “Proceed To Simulation Results” button in the pop-up box to track the simulation processing status and simulation results.

![Figure 2.8: “Launch Successful” message after launching simulation in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.](image)
2.2.3 Simulation Results

In the Simulation Results tab, users can view the simulation progress and simulation results (§2.2.3.1), restore the simulation settings if needed (§2.2.3.2), and download intelligent simulation reports (§2.2.3.3). Specifically, all the simulation results (figures and tables) can be downloaded in Word format. Hereinafter, we use simulation results and operating characteristics interchangeably.

2.2.3.1 View simulation results

In the Simulation Results tab, the Running Simulations panel exhibits the progress of ongoing simulation (Figure 2.9). The ongoing simulations are displayed in ascending order by the launch time. Click the icon “×” to delete the corresponding simulation.

![Running Simulations panel](image)

**Figure 2.9:** Simulation progress in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

Once the simulations are completed, the Running Simulations panel in Figure 2.9 will disappear, a green “simulation result created” massages will appear instead and stay at the same place of the Running Simulations panel unless explicitly dismissed by clicking the icon “×” at the end of the corresponding row, and the simulation results will be automatically loaded into the Simulation History panel (Figure 2.10), with the blue mail icon shown to indicate new results. All the previously completed simulations are also listed in the Simulation History panel. Simulation results for other modules can also be viewed under the Simulation History by dropping down the “Select a module” button (Figure 2.10). Click the button to delete the selected simulation results.
Module 2. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment

Figure 2.10: Simulation Results in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

Click the button to unfold the simulation results (Figure 2.10). The design settings are firstly displayed at the top of each simulation study (Figure 2.11). Then the results of simulation are shown as plots and tables below.

Figure 2.11: View the simulation results in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.
Simulation Result Plots

There are two sections in the Simulation Result Plots:

A. Line plots showing six summary statistics of the simulation results for all the designs (Figure 2.12), including **Prob. of Selecting MTD**, **Prob. of Toxicity**, **Prob. of Selecting Does-over-MTD**, **Prob. of No Selection**, **Trial Duration** and **Average Number of Enrolled Patients**. 

B. A table of mean and standard deviation (s.d.) for the six summary statistics (Figure 2.13).

A. Line plots:

- The six summary statistics are part of operating characteristics of the designs. They are explained in full detail next.

  - **Prob. of Selecting MTD**: The probability of selecting the true MTD, defined as the proportion of simulated trials that correctly select the true MTD. The higher the value, the better the design.

    * For interval-based designs (mTPI-2 & R-TPI), the true MTDs are defined as the dose levels of which the true toxicity probabilities fall into the equivalence interval \([p_T - \epsilon_1, p_T + \epsilon_2]\); if none of the doses have a toxicity probability that falls into the equivalence interval, the true MTD is defined as the dose with the highest toxicity probability below \(p_T\). For the non-interval-based designs, 3+3 and Rolling 6, the true MTDs is defined as the dose levels with the highest toxicity probabilities lower than or equal to \(p_T\).

    * To compare the operating characteristics of multiple designs submitted in a simulation study, the definition of MTD should be unified. If any of interval-based designs (mTPI-2 & R-TPI) are used in the simulation, the dose levels of which the true toxicity probabilities fall into the widest equivalence interval \([p_T - \max\{\epsilon_1\}, p_T + \max\{\epsilon_2\}]\) are defined as the true MTDs. Here, \(\max\{\cdot\}\) is taken over the designs. If none of the doses fall in, the dose with the highest toxicity probability that is below \(p_T\) is the true MTD. For example, consider a case in which users compare four designs, R-TPI, mTPI-2, Rolling 6 and 3+3, in a simulation study targeting \(p_T = 0.3\). Suppose \(\epsilon_1 = 0.02\) and \(\epsilon_2 = 0.05\) for R-TPI, and \(\epsilon_1 = 0.05\) and \(\epsilon_2 = 0.03\) for mTPI-2. In this case, the true MTD is the dose levels with toxicity probabilities in \([0.3 - 0.05, 0.3 + 0.05]\); if none of the doses have a toxicity probability in \([0.3-0.05, 0.3+0.05]\), the dose with the highest toxicity probability lower than 0.3 is the true MTD.

    * If a scenario does not have any MTD (e.g., all doses have toxicity probabilities...
larger than the target \( p_T \), no selection is the right decision. In this case, the probability of selecting the true MTD is the probability of no selection.

- **Prob. of Toxicity**: The proportion of patients who have experienced DLT across all the simulated trials. The lower the number, the fewer patients having DLTs under the design.

- **Prob. of Selecting Does-over-MTD**: The probability of selecting the dose levels above the true MTD, which is defined by the proportion of simulated trials that select a dose higher than the true MTD at the end of the trial. The lower the value, the better the safety of the design.

- **Prob. of No Selection**: The proportion of the simulated trials in which none of the dose levels are selected as the MTD. If a scenario does not have any MTD, this values is treated as the probability of selecting the true MTD.

- **Average Trail Duration**: The average time duration for trial conduct (in days). The lower the value, the faster the trials and the more economic of the design.

- **Average Number of Enrolled Patients**: The average number of patients enrolled in the trial, including the patients who complete the DLT observation period with DLT or non-DLT, and patients who drop out of the trial and become inevaluable for DLTs.

  - For each line plot, the x-axis is the index of scenario and the y-axis is the value of summary statistics. Lines with different colors represent different designs.

  - The plots are interactive for better visualization.
    - Hover the mouse on a dot and a box will display the value of each design at the corresponding scenario (e.g. top left plot in Figure 2.12: Prob. of Selecting MTD)
    - Hover the mouse on the design label to highlight the corresponding line and fade the others (e.g. bottom right plot in Figure 2.12: Average Number of Enrolled Patients).
    - Click the design label to hide the corresponding line and click again to change it back (e.g. bottom right plot in Figure 2.12: Average Number of Enrolled Patients).

B. Simulation summary table: Figure 2.13 shows the mean±sd of the summary statistics across all scenarios for each design.
2.2. User Interface and Tutorial
2.2.3. Simulation Results

Figure 2.12: Simulation result plots in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

Figure 2.13: Simulation summary table in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

Summary of Performance

<table>
<thead>
<tr>
<th></th>
<th>Design 1 (3+3)</th>
<th>Design 2 (mTP1-2)</th>
<th>Design 3 (Rolling 6)</th>
<th>Design 4 (R-TP1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob. of Select MTD</td>
<td>0.265 ± 0.086</td>
<td>0.391 ± 0.091</td>
<td>0.239 ± 0.067</td>
<td>0.423 ± 0.103</td>
</tr>
<tr>
<td>Prob. of Toxicity</td>
<td>0.222 ± 0.074</td>
<td>0.211 ± 0.065</td>
<td>0.229 ± 0.076</td>
<td>0.222 ± 0.067</td>
</tr>
<tr>
<td>Prob. of Select Dose-over-MTD</td>
<td>0.085 ± 0.064</td>
<td>0.153 ± 0.126</td>
<td>0.065 ± 0.061</td>
<td>0.126 ± 0.092</td>
</tr>
<tr>
<td>Prob. of No Selection</td>
<td>0.158 ± 0.183</td>
<td>0.02 ± 0.039</td>
<td>0.179 ± 0.209</td>
<td>0.029 ± 0.052</td>
</tr>
<tr>
<td>Trial duration</td>
<td>242.154 ± 73.562</td>
<td>265.769 ± 76.185</td>
<td>204.228 ± 72.125</td>
<td>375.208 ± 13.787</td>
</tr>
<tr>
<td>Number of enrolled patients</td>
<td>15.565 ± 4.342</td>
<td>17.095 ± 4.537</td>
<td>17.35 ± 5.956</td>
<td>29.353 ± 1.131</td>
</tr>
</tbody>
</table>

* Mean ± Standard deviation. The statistics are calculated given by the current scenario and design setting.
Simulation Result Tables

Full simulation results are presented in tabular format arranged by scenarios (Figure 2.14).

In the upper part of Figure 2.14, the first two columns summarize dose levels and their true toxicity probabilities; the remaining columns report three dose-specific summary statistics from the simulations: selection probability, average number of patients treated, and average number of toxicities (i.e. DLTs), along with their standard deviations, at each dose level. Specifically, they are

1) Selection Prob.: The proportion of simulated trials that select each dose level as the MTD.
2) # of Patients Treated (s.d.): The average number of patients treated at each dose level.
3) # of Toxicities (s.d.): The average number of patients experienced DLT at each dose level.

The true MTD(s) of the scenario is(are) highlighted by the orange bar. For the definition of the true MTD in the simulation results, please refer to the definition of Prob. of Selecting MTD in the Simulation Results Plots above (after Figure 2.11).

In the lower part of Figure 2.14, more trial-specific summary statistics are reported, mainly from four aspects: MTD Selection, Trial Toxicity, Trial Duration and Trial Sample Size.

- MTD Selection
  - Prob. of Selecting MTD: The proportion of simulated trials that select the true MTD at the end of the trial.
  - Prob. of Selecting Does-over-MTD: The proportion of simulated trials that select the doses higher than the true MTD at the end of the trial.
  - Prob. of No Selection: The proportion of simulated trials in which none of the dose levels are selected as the MTD. If a scenario does not have any MTD, this values is treated as the probability of selecting the true MTD.

For detailed descriptions, please refer to Simulation Result Plots section above (after Figure 2.11).

- Trial Toxicity
  - Prob. of Toxicity: The proportion of patients experiencing DLT across all the simulated trials. For detailed descriptions, please refer to Simulation Result Plots section above (after Figure 2.11).

- Trial Duration
  - Average Trial Duration (s.d.): The average time duration for trial conduct (in days) and its standard deviation.

- Trial Sample Size
  - Average # of Patients Treated (s.d.): The average number of patients treated in the
2.2. User Interface and Tutorial
2.2.3. Simulation Results

Simulated trials and its standard deviation. Due to early stopping, this number is lower than or equal to \( n \).

When calculating the standard deviation, we use \( n_{\text{sim}} \) as the denominator instead of \( (n_{\text{sim}} - 1) \) in U-Design.

### Scenario 1

<table>
<thead>
<tr>
<th>Dose Level</th>
<th>True Pos Prob.</th>
<th>Selection Prob.</th>
<th># of Patients Treated</th>
<th># of Toxicities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3=5</td>
<td>mTIP-2</td>
<td>Rolling 6</td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.925</td>
<td>0.405</td>
<td>4.917</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
<td>0.283</td>
<td>0.15</td>
<td>4.966</td>
</tr>
<tr>
<td>3</td>
<td>0.45</td>
<td>0.067</td>
<td>0.022</td>
<td>0.905</td>
</tr>
<tr>
<td>4</td>
<td>0.75</td>
<td>0.009</td>
<td>0.001</td>
<td>0.004</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.001</td>
<td>0</td>
<td>0.001</td>
</tr>
</tbody>
</table>

When calculating the standard deviation, we use \( n_{\text{sim}} \) as the denominator instead of \( (n_{\text{sim}} - 1) \) in U-Design.

### Scenario 2

<table>
<thead>
<tr>
<th>Dose Level</th>
<th>True Pos Prob.</th>
<th>Selection Prob.</th>
<th># of Patients Treated</th>
<th># of Toxicities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3=5</td>
<td>mTIP-2</td>
<td>Rolling 6</td>
</tr>
<tr>
<td>1</td>
<td>0.08</td>
<td>0.222</td>
<td>0.068</td>
<td>4.126</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>0.131</td>
<td>0.32</td>
<td>0.899</td>
</tr>
<tr>
<td>3</td>
<td>0.24</td>
<td>0.288</td>
<td>0.243</td>
<td>0.906</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>0.125</td>
<td>0.262</td>
<td>0.867</td>
</tr>
<tr>
<td>5</td>
<td>0.38</td>
<td>0.076</td>
<td>0.065</td>
<td>1.039</td>
</tr>
</tbody>
</table>

The row with ** background color indicates the TRUE MTD.

**Figure 2.14:** Simulation result tables in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.
2.2.3.2 Restore simulation

Users can restore the simulation settings from the simulation results by clicking the “Restore” button at the upper right corner of each simulation results panel (yellow arrow in Figure 2.15). Upon clicking, the display will switch to the Simulation Setup page with the same simulation settings restored. This is useful to restore the old simulation settings for reproducible results.

Figure 2.15: Restore simulation setup and download simulation results in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

2.2.3.3 Download simulation results

There is a “Download Report” button at the upper right corner of each simulation results panel (green arrow in Figure 2.15). Click it to download a word file, which includes three parts:

- Part A: Complete simulation results under the designs and scenarios users added in the Simulation Setup tab;
- Part B: Detailed technical descriptions of the designs users added in the Simulation Setup tab;
- Part C: Reference

Users may select the required parts and modify them tailored for their trials or contact us via email (admin@laiyaconsulting.com) for consulting services.
2.2.4 Decision Table

In the Decision Table tab, users can generate decision tables of mTPI-2 and 3+3 designs, to guide the dose escalation/de-escalation during trial conduct. The decision tables of Rolling 6 and R-TPI can also be pretabulated, but are much more complicated, so U-Design do not provide them in current version.

Manually type in the maximum number of patients at a dose \( n \), target toxicity probability \( p_T \) and two small fractions \( \epsilon_1 \) and \( \epsilon_2 \) for decision table generation (Figure 2.17). Hover mouse over each parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 2.4.

![Image](image.png)

**Figure 2.16:** Input parameters in the Decision Table tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

Click the “Generate” button to generate five decision tables for two different designs at the same time (Figure 2.17). Users can click the tabs to switch between the tables for the mTPI-2 and 3+3 designs.

Click the “Download Decision Table” button to save the decision table of the corresponding design in word (.docx).

For each decision table, the column represents the number of patients treated at a dose, which is mostly used for the current dose, the dose currently being used to treat patients in the trial, and the row represents the number of patients among those treated at that dose who have experienced dose-limiting toxicity (DLT) events. Note that these are the counts of patients, not DLT events. For example, column 3 and row 1 means that 3 patients have been treated at the current dose and 1 of them experiences DLT. Each cell in the decision table provides the dose-assignment decision based on the readouts from the corresponding row and column. For example, for column 3 and row 1, i.e., 1 out of 3 patients experiences DLTs, the decision is “S”. The letters in the decision table represent different dose-assignment decisions as shown below:
Table 2.4: Input arguments in the Decision Table tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Number of patients at a dose</td>
<td>The maximum number of patients to be treated at a dose. Here, the upper limit is set at 30 since the number of patients that are enrolled at a dose in phase I clinical trial is typically small.</td>
</tr>
<tr>
<td>$p_T$</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). The main objective of phase I clinical trials is to find the highest dose with a toxicity probability closest to or lower than $p_T$.</td>
</tr>
<tr>
<td>$\epsilon_1, \epsilon_2$</td>
<td>$\epsilon_1$: lower margin $\epsilon_2$: higher margin</td>
<td>Two small fractions used to define the equivalence/target interval of the MTD. Any dose with a toxicity probability falling into the interval $[p_T - \epsilon_1, p_T + \epsilon_2]$ is considered an acceptable dose MTD. Default values for both are 0.05.</td>
</tr>
</tbody>
</table>

- “E” stands for escalating to the next higher dose,
- “S” stands for staying at the current dose,
- “D” stands for de-escalating to the previous lower dose,
- “DU” stands for de-escalating to the previous lower dose and marking the current dose and its higher doses as unacceptably toxic so that they will never be used again in the remainder of the trial.

The 3+3 decision table is fixed regardless of different trial parameters. For rolling six and R-TPI designs, their decisions are not only based on the number of patients treated at a dose and the number of patients having already experienced DLT events among them, but also on the number of patients who are still being followed without outcomes. See the decision table of the rolling six in Table 2.6 (§2.3.4). See an example of R-TPI decision table in Table 2.8 (§2.3.5).
Figure 2.17: Decision tables generated in the **Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment** module.
Module 2. Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment

2.2.5 MTD Estimation

In the MTD Estimation tab, users can estimate the MTD for mTPI-2 and R-TPI designs based on the isotonic regression through Pool Adjacent Violators Algorithm (PAVA), after the dose finding is completed and the DLT outcomes of all patients are collected.

Specify the target toxicity probability ($p_T$), and two small fractions to define the equivalence interval ($\epsilon_1$ and $\epsilon_2$) in the design. Select the number of doses ($n_{dose}$) from the dropdown box, then an editable table will be shown below on the page (Figure 2.18). Then manually type in the observed number of toxicities (DLTs) and the number of patients treated at each dose into the table and click the “Estimate” button to estimate the MTD. Finally, the estimated MTD is highlighted in blue background as shown in Figure 2.19.

Hover mouse over each parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 2.5.

![Figure 2.18: Input parameters in the MTD Estimation tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.]

![Figure 2.19: MTD estimation in the Single-Agent Dose-Finding Designs with Toxicity Endpoint and Rolling Enrollment module.]

Your trial designs anywhere, anytime.
Table 2.5: Input parameters in the MTD Estimation tab of Single-Agent Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T$</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). The main objective of phase I clinical trials is to find the highest dose with a toxicity probability closest to or lower than $p_T$.</td>
</tr>
<tr>
<td>$\epsilon_1, \epsilon_2$</td>
<td>$\epsilon_1$: lower margin $\epsilon_2$: higher margin</td>
<td>Two small fractions used to define the equivalence/target interval of the MTD. Any dose with a toxicity probability falling into the interval $[p_T - \epsilon_1, p_T + \epsilon_2]$ is considered an acceptable dose MTD. Default values for both are 0.05.</td>
</tr>
<tr>
<td>$n_{dose}$</td>
<td>The number of doses</td>
<td>The number of candidate dose levels for investigation</td>
</tr>
<tr>
<td># of DLTs</td>
<td>The number of patients with DLTs at each dose level</td>
<td>A non-negative integer number of patients with DLT at each dose level</td>
</tr>
<tr>
<td># of patients</td>
<td>The number of patients treated at each dose level</td>
<td>A positive integer number of patients treated at each dose level, which should be no less than the # of DLTs</td>
</tr>
</tbody>
</table>
2.3 Statistical Methods Review

2.3.1 Simulating Patients Enrollment and Evaluation

To better demonstrate the benefit of rolling-based designs in accelerating the trial conduct, trial duration would be assessed, in addition to the operating characteristics of the safety and reliability. Therefore, this module simulates trials based on a real-life setting, in order to better reflect the real-world situation. Figure 2.20 illustrates the simulation process of patients enrollment and evaluation. Specifically,

1. Each patient is assigned an inter-patient arrival time (a chronological time for a patient to arrive in the clinic and be eligible for study). The inter-patient arrival time is sampled from gamma distribution, with the shape parameter of $a$ and scale parameter of $b$. So the mean inter-patient arrival time (MIAT) is $\text{MIAT} = ab$. For example, MIAT is 10 or 5 days means on average three or six patients per month will arrive in the clinic and be eligible for study, respectively.

2. To mimic real-life oncology dose-finding trials, each enrolled patient in the simulation study is also assigned an on-study start time (the gap between the time of arrival in the clinic and the starting time of treatment) and the probability of inevaluability (such as drop off).
   (a) A random binary DLT/non-DLT outcome generated with the true probability of toxicity for the corresponding dose at which the patient is assigned.
   (b) A random binary evaluability/inevaluability outcome generated with the inevaluable rate (IR) of the enrolled patient.
   (c) The on-study start time is sampled from the uniform distribution ranging from 0 to the maximum waiting time.
   (d) If a DLT occurs in that patient, the time to DLT is sampled from the uniform distribution ranging from 0 to the maximum DLT follow-up period $T_{\text{follow-up}}$; Otherwise, the time to non-DLT is $T_{\text{follow-up}}$.
   (e) If that patient becomes inevaluable, the time to inevaluability (IE) of that patient is sampled from a uniform distribution ranging from 0 to the sampled time to event (either DLT or non-DLT) of that patient.

Therefore, assume that the trial starts at the time $t = 0$ (i.e., the first patient arrives and is available for study at the time $t = 0$), a patient complete the trial with the event (DLT, non-DLT, or IE) at the time $t_i = \text{The arrival time} + \text{The on-study start time} + \text{The time to DLT or non-DLT or IE}$.
2.3. Statistical Methods Review

2.3.1. Simulating Patients Enrollment and Evaluation

In U-Design, we fix \( a = 1 \) in sampling the inter-patient arrival time, so the mean inter-patient arrival time is \( \text{MIAT} = b \). And also, for simplicity, we assume that there is no waiting time between the time of patient arrival in the clinic and the starting time of treatment, so the on-study start time is 0.

**Figure 2.20**: Simulating patients enrollment and evaluation in the Single Agent – Rolling-Based Designs.
2.3.2 The 3+3 Design

The 3+3 design (Storer, 1989) is a rule-based design which enrolls patients in a cohort of three. It starts by allocating the first cohort of three patients to the starting dose (which is often the lowest dose level) and adaptively escalates/de-escalates to the next dose level based on observed number of dose limiting toxicities (DLTs). Besides, in this module, the ethics constraint of “decision-in-advance” (§2.3.2.2) is adopted, which is applicable to the real-life trials.

2.3.2.1 Design Algorithm

In 3+3, a maximum of six patients are allowed to be treated at any dose level, and the MTD is defined as the highest dose for which one or fewer DLTs occurred in six patients. Its algorithm proceeds as follows:

0. Start the trial by treating three patients at a prespecified starting dose level.
1. Escalate to the next higher dose or de-escalate to the previous lower dose according to the following rules:
   (a) If 0 of 3 patients has a DLT, escalate to next higher dose and treat three patients.
   (b) If 2 or more of 3 patients have DLTs, de-escalate to previous lower dose and treat three patients.
   (c) If 1 of 3 patients has a DLT, treat three more patients at current dose level.
      i. If 1 of 6 has DLT, escalate to next higher dose and treat three patients if the next higher dose has not been tried; otherwise, declare it as the MTD and stop the trial.
      ii. If 2 or more of 6 have DLTs, de-escalate to previous lower dose level and treat three patients.
   (d) If the trial de-escalates to previous lower dose:
      i. If only 3 or less had been treated at the previous lower dose, treat three more patients at that dose.
      ii. If six have already been treated at the previous lower dose, stop the trial and declare the lower dose as the MTD.
2. Escalation never occurs to a dose at which two or more DLTs have already occurred.
3. If de-escalation occurs at the lowest dose, the trial is stopped.
4. Repeat steps 1-3 until either the MTD is identified or the trial is stopped for excessive toxicity.
2.3.2.2 The “Decision-in-Advance” Rule

If a cohort of patients for the current dose is not fully enrolled or completely observed, a decision can be made in advance if and only if the decision would not be changed by the pending data for the cohort of patients, either enrolled but still being followed or yet to be enrolled. For example, under 3+3, if 2 patients have been enrolled to a newly used dose \( d \) and both of them experience DLTs, stop enrolling the third patient to \( d \) and de-escalate to \( d - 1 \) immediately. This rule of “decision-in-advance” can accelerate the trial and make the trial conduct more ethical.
2.3.3 The Modified Toxicity Probability Interval-2 (mTPI-2) Design

The modified toxicity probability interval-2 (mTPI-2) design (Guo et al., 2017) is a cohort-based design which enrolls patients according to a pre-planned cohort size. It is also a model-base design, which uses a simple beta-binomial model to estimate the toxicity probability and makes dose escalation/de-escalation decisions based on the unit probability mass (UPM) of a series of dosing interval with equal length. At the end, mTPI-2 selects the dose of which the isotonic transformed toxicity probability is the closest to the target \( p_T \) as the MTD. Besides, in this module, the ethics constraint of “decision-in-advance” (§2.3.3.3) is adopted, which is applicable to the real-life trials.

2.3.3.1 Probability Model

Consider a phase I trial with \( D \) candidate doses for escalation. Let \( p_1, \ldots, p_D \) denote the true toxicity probabilities for doses \( d = 1, \ldots, D \). The observed data include \( n_d \), the number of patients treated at dose \( d \), and \( y_d \), the number of patients experiencing a toxicity. Let \( Data = \{(y_d, n_d); d = 1, 2, \ldots, D\} \).

The mTPI-2 design employs a simple beta-binomial hierarchical model as follow:

\[
\begin{align*}
    y_d \mid n_d, p_d &\sim binomial(n_d, p_d) \\
p_d &\sim beta(\alpha, \beta)
\end{align*}
\]

The posterior distribution of \( p_d \) is given by

\[
p_d \mid y_d, n_d \sim beta(\alpha + y_d, \beta + n_d - y_d).
\]

In U-Design, we adopt the prior \( beta(1, 1) \) for \( p_d \), because it would lead to slightly conservative posterior inference as the prior mean is 0.5, which is usually above \( p_T \).

2.3.3.2 Dose-Finding Rules

Equal-width Dosing Intervals: The mTPI-2 design improves over the mTPI design (Ji et al., 2010) by blunting the Ockhams razor that leads to some statistically sound but practically debatable decisions in mTPI. In mTPI, the unit interval \((0, 1)\) is divided into three intervals: the under-dosing interval \((0, p_T - \epsilon_1)\), the equivalence interval \((p_T - \epsilon_1, p_T + \epsilon_2)\), and the over-dosing interval \((p_T + \epsilon_2, 1)\). Here, \( \epsilon_1 \) and \( \epsilon_2 \) are small fractions, such as 0.05, to account for the uncertainty around the true target toxicity \( p_T \). However, mTPI-2 resolves the the Ockhams razor problem fundamentally by dividing the intervals \((0, p_T - \epsilon_1)\) and \((p_T + \epsilon_2, 1)\) into shorter subintervals with length \((\epsilon_1 + \epsilon_2)\),
which is the same as the length of the equivalence interval, to mitigate the effect of interval length in the mTPI design. For clarify, denote $EI$ the equivalence interval $(p_T - \epsilon_1, p_T + \epsilon_2)$, and $LI$ a set of intervals below $EI$,

$$LI = \{M_{LI}^1 = (p_T - 2\epsilon_1 - \epsilon_2, p_T - \epsilon_1), M_{LI}^2 = (p_T - 3\epsilon_1 - 2\epsilon_2, p_T - 2\epsilon_1 - \epsilon_2), \cdots, M_{LI}^J = (0, p_T - J\epsilon_1 - (J-1)\epsilon_2)\}$$

and $HI$ a set of intervals above $EI$,

$$HI = \{M_{HI}^1 = (p_T + \epsilon_2, p_T + \epsilon_1 + 2\epsilon_2), M_{HI}^2 = (p_T + \epsilon_1 + 2\epsilon_2, p_T + 2\epsilon_1 + 3\epsilon_2), \cdots, M_{HI}^K = (p_T + (K-1)\epsilon_1 + K\epsilon_2, 1)\}$$

For example, if $p_T = 0.3$ and $\epsilon_1 = \epsilon_2 = 0.05$,

$$EI = (0.25, 0.35)$$

$$LI = \{M_{LI}^1 = (0.15, 0.25), M_{LI}^2 = (0.05, 0.15), M_{LI}^3 = (0, 0.05)\}$$

$$HI = \{M_{HI}^1 = (0.35, 0.45), M_{HI}^2 = (0.45, 0.55), M_{HI}^3 = (0.55, 0.65), M_{HI}^4 = (0.65, 0.75), M_{HI}^5 = (0.75, 0.85), M_{HI}^6 = (0.85, 0.95), M_{HI}^7 = (0.95, 1)\}$$

Other than the boundaries $(0, 0.05)$ and $(0.95, 1)$, all the intervals have the same length. The boundaries do not affect the decision making since they are clearly associated with “E” and “D” decisions, respectively. See Guo et al. (2017) for details.

**Dose-Finding Rules:** Given the interval and a probability distribution like (2.1), define the unit probability mass (UPM) of that interval as the probability of the interval divided by the length of the interval. Mathematically, the UPM of an interval $(a, b)$ equals to

$$\text{UPM} = \frac{\text{Prob}\{p \in (a, b) \mid \text{Data}\}}{b - a}$$

The mTPI-2 selects the (sub-)interval with the largest UPM value as the winning interval and take the dose-escalation decision corresponding to the winning (sub-)interval. More specifically,

- If the equivalence interval $M^{EI} = (p_T - \epsilon_1, p_T + \epsilon_2)$ has the largest UPM, it is selected as the winning interval and dose-assignment decision of mTPI-2 is “S”, to stay at the current dose.
- If any interval $M_{LI}^J$ in $LI$ has the largest UPM, it is selected as the winning interval and dose-assignment decision of mTPI-2 is “E”, to escalate to the next higher dose.
- If any interval $M_{HI}^K$ in $HI$ has the largest UPM, it is selected as the winning interval and dose-assignment decision of mTPI-2 is “D”, to de-escalate to the previous lower dose.

*Your trial designs anywhere, anytime.*
2.3.3.3 The “Decision-in-Advance” Rule

If a cohort for the current dose is not fully enrolled or completely observed, a decision can be made in advance if and only if the decision would not be changed by the pending data for the cohort of patients, either enrolled but still being followed or yet to be enrolled. For example, under $p_T = 0.3$, if 2 patients have been enrolled to a newly-used dose $d$ and both of them experience DLTs, stop enrolling the third patient to $d$ and de-escalate to $d - 1$ immediately. This is ethical and sensible, as the decision would still be de-escalation if a third patient is enrolled to dose $d$ and experiences non-DLT eventually. This rule of “decision-in-advance” can accelerate the trial and make the trial conduct more ethical.

2.3.3.4 Safety Rules

For trial safety, two additional rules are applied.

- [Rule 1: Dose Exclusion] If the current dose is considered excessively toxic, i.e., $\text{Prob}\{p_d > p_T \mid \text{Data}\} > \xi$, where the threshold $\xi$ is close to 1, say 0.95, the current and all higher doses are excluded and never be used again in the remainder of the trial.

- [Rule 2: Early Stop] If the current dose is the lowest dose (first dose) and is considered excessively toxic according to Rule 1, early stop the trial for safety.

In safety Rules 1 and 2, $\text{Prob}\{p_d > p_T \mid \text{Data}\}$ is a function of the cumulative beta distribution in (2.1).

2.3.3.5 Trial Termination

The trial proceeds until any of the following conditions is satisfied:

1. If the prespecified maximum total sample size is reached;
2. If the lowest dose shows excessive toxicity according to Rule 2; In this case, the trial is early stopped and the MTD cannot be determined;
3. Optional: ad-hoc rules of maximum number of patients at a dose, denoted by $K (K < n)$:
   - If the mTPI-2 decision is “S”, to stay at the current dose, and the current dose has enrolled $K$ patients;
   - If the mTPI-2 decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled $K$ patients;
   - If the mTPI-2 decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled $K$ patients.
2.3.3. The Modified Toxicity Probability Interval-2 (mTPI-2) Design

2.3.3.6 MTD Selection

Once all the enrolled patients complete the DLT observation and the trial is not stopped early, the mTPI-2 design applies an isotonic regression to select the MTD. Follow the steps as below.

1. Compute the isotonically transformed posterior means of DLT probabilities for all the dose levels in the following two steps.
   (a) Using the accumulated safety information about $y_d$ and $n_d$ for $d = 1, \ldots, D$, compute the posterior mean and variance for all the dose levels, \{\tilde{p}_1, \ldots, \tilde{p}_D\} and \{v_1, \ldots, v_D\}. Here in U-Design, an independent prior $\text{beta}(0.005, 0.005)$ is used to compute the posterior mean and variance.
   (b) Compute isotonic regression estimates of the posterior mean by solving the optimization problem, minimizing $\sum_{d=1}^{D} (\hat{p}_d - \tilde{p}_d)^2 / v_d$ subject to $\hat{p}_j \geq \hat{p}_k$, for $j > k$. Such optimization can be done using the pooled adjacent violators algorithm (PAVA), the estimated DLT probabilities satisfying the order constraint is obtained, denoted by \{\hat{p}_1, \ldots, \hat{p}_D\}.

2. Among all the tried doses for which $\text{Prob}\{p_d > p_T \mid Data\} < \xi$ and $\hat{p}_d \leq p_T + \epsilon_2$, select as the estimated MTD the dose with the smallest difference $|\hat{p}_d - p_T|$. That is, the estimated MTD is $d^* = \text{argmax}_d |\hat{p}_d - p_T|$.

3. In case of a tie (i.e., two or more doses have the smallest difference),
   (a) If there is at least one dose lower the target $p_T$ among all the tied doses, choose the highest dose among those as the estimated MTD;
   (b) Otherwise, choose the lowest dose among the tied doses as the estimated MTD.
2.3.4 The Rolling Six Design (RSD)

The rolling six design (RSD) (Skolnik et al., 2008) extends 3+3 with the aim to reduce the occurrence of accrual suspension after enrolling each three patients and accelerate the trial conduct. It allows for accrual of two to six patients concurrently onto a dose level based on the number of patients concurrently enrolled and evaluable (# Enrolled), the number experiencing dose-limiting toxicity (DLT) (# DLTs), and the number still at risk of developing a DLT (# Pending). The rolling six is a rule-based design and all dose assignment rules for the six patients can be pretabulated (see Table 2.6).
### Table 2.6: The decision table of the rolling six design.

<table>
<thead>
<tr>
<th># Enrolled</th>
<th>Observed data at dose ( d )</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># DLTs</td>
<td># Non-DLTs</td>
</tr>
<tr>
<td>2</td>
<td>0, 1</td>
<td>any</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0, 1, 2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0, 1, 2</td>
</tr>
<tr>
<td>3</td>
<td>( \geq 2 )</td>
<td>any</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>4</td>
<td>( \geq 2 )</td>
<td>any</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0, 1, 2, 3, 4</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0, 1, 2, 3, 4</td>
</tr>
<tr>
<td>5</td>
<td>( \geq 2 )</td>
<td>any</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0, 1, 2, 3, 4</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>5, 6</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0, 1, 2, 3, 4</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>( \geq 2 )</td>
<td>any</td>
</tr>
</tbody>
</table>

**NOTE.** 1) This table does not take into account inevaluable patients, such as patients who drop off during the DLT observation period; 2) Escalation never occurs to a dose at which 2 or more DLTs have already occurred, because that dose level is considered excessive toxicity and should be excluded in the remaining dose finding; 3) If de-escalation occurs at the lowest dose level, then the study is early terminated.

**Abbreviations:** DLT, dose-limiting toxicity; MTD, maximum tolerated dose; E, escalate to next higher dose level; S, stay at the current dose level; D, de-escalate to previous lower dose level.
2.3.5 The Rolling Toxicity Probability Interval (R-TPI) Design

This section describes the rolling toxicity probability interval (R-TPI) design proposed by Guo et al. (2019). R-TPI design combines the idea of rolling accrual in rolling six design (Skolnik et al., 2008) with the model-based framework in mTPI-2 (Guo et al., 2017). Accordingly R-TPI enjoys the benefits of model-based inference and overcomes the drawback of fixed cohort size.

2.3.5.1 Notations

Consider a toxicity-driven phase I dose-finding trial. Let $p_T$ be the target DLT probability, and $p_d$ be the true and unknown DLT probabilities of dose level $d$, $d = 1, \ldots, D$, where $D$ denotes the prespecified number of dose levels to be investigated. Generally, we assume that $p_d$ is non-decreasing with dose level, i.e. $p_1 \leq p_2 \leq \cdots \leq p_D$. Assume at a given moment, dose $d$ is being used to treat enrolled patients and a total of $(n_d + m_d)$ patients have been assigned to dose $d$, among whom $n_d$ patients have known outcomes (either with or without DLT) and $m_d$ patients are still being followed without outcomes. Let $y_d$ be the number of patients (among $n_d$) with DLT, therefore $(n_d - y_d)$ without DLT. The table below describes the breakdowns.

<table>
<thead>
<tr>
<th># with DLT</th>
<th># without DLT</th>
<th># being followed and no outcomes</th>
<th>Total at dose $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_d$</td>
<td>$(n_d - y_d)$</td>
<td>$m_d$</td>
<td>$(n_d + m_d)$</td>
</tr>
</tbody>
</table>

2.3.5.2 Dose-Finding Rules and Design Algorithm

The R-TPI design consists of two sets of enrollment schemes, namely the run-in enrollment and the rolling enrollment. To begin the trial, R-TPI enrolls the first patient at the starting dose level.

Run-in Enrollment The run-in enrollment is applied to any new dose level when it is first used to treat patients during the trial. Suppose dose $d$ is decided to be the new dose level for treating patients and it has not been used at any time of the trial. R-TPI starts run-in enrollment and keeps enrolling new patients at dose $d$ until any one of two below case,

1. $n_d > 0$, i.e. there is at least one outcome at $d$,
2. $n_d = 0$ and $m_d = C$, for a pre-determined $C$ value, i.e., the first $C$ patients have not completed follow-up at $d$ and are without definitive outcomes. Here, $C$ is the maximum number of pending patients without observed outcomes allowed in the trial such that a new patient can be enrolled. For example, for the rolling six design, $C = 6$.

And then,

- in the case (1), R-TPI starts rolling enrollment (specified below).
2.3. Statistical Methods Review

2.3.5. The Rolling Toxicity Probability Interval (R-TPI) Design

– in the case (2), R-TPI first suspends the enrollment until the first outcome at current dose \(d\) and then starts the rolling enrollment (specified below).

**Rolling Enrollment** Supposes at a given moment of the trial a new patient becomes eligible for enrollment, and the current dose used for treating patients is \(d\) at which \((n_d + m_d)\) patients have been treated. To explain the rolling enrollment, there are two more notations needed to be introduced.

- \(k_d\): number of patients at dose \(d\) since it most recently becomes the current dose.
  
  For example, if initially three patients are enrolled at dose level \(d\), and based on their DLT outcomes R-TPI changes the dose level to another dose and enrolls patients at the new dose; however, based on the patients DLT outcomes at the new dose R-TPI changes the dose level again, switches back to dose \(d\), and enrolls additional 3 patients. At this time \(k_d = 3\).

- \(D_{y_d, n_d}\): the mTPI-2 decision based on the toxicity data of \(y_d\) out of \(n_d\) patients experiencing DLTs at dose \(d\), \(D_{y_d, n_d} \in \{D, E, S\}\). Here, “D” stands for de-escalating to the previous lower dose level \(d - 1\), “E” for escalating to the next higher dose level \(d + 1\), and “S” for staying at the current dose level \(d\). For the detailed mTPI-2 dose escalation rule, please refer to §2.3.3.2.

The dose-assignment decisions in the rolling enrollment is mainly based on the mTPI-2 decision of current observation \(D_{y_d, n_d}\), the mTPI-2 decision of the most toxic scenario where all pend-

| Table 2.7: The R-TPI dose-finding rules applied in the rolling enrollment. |
|------------------|------------------|------------------|------------------|------------------|
|                  | mTPI-2 decision for | mTPI-2 decision for | mTPI-2 decision for | R-TPI Decision   |
|                  | current observation | the most toxic scenario | the safest scenario |                  |
| \((D_{y_d, n_d})\)| \((D_{y_d+m_d, n_d+m_d})\) | \((D_{y_d, n_d+m_d})\) |                  |
| Case 1           | D                | D                | D                | D                |
| Case 2           | D                | D                | S or E            | S                |
| Case 3           | S                | S or D           | S                | S                |
| Case 4           | S                | S or D           | E                | S or Suspend*    |
| Case 5           | E                | E                | E                | E                |
| Case 6           | E                | S or D           | E                | S or Suspend*    |

* If 3 or more continuous patients has been enrolled to the same dose \((k_d > 3)\), suspend the trial to avoid over-enrolling patients on the current dose.

Abbreviations: E, escalate to next higher dose level; S, stay at the current dose level; D, de-escalate to previous lower dose level.
ing patients experience DLTs in the future $D_{y_d+m_d,n_d+m_d}$, and the mTPI-2 decision of the safest scenario where none of pending patients experiences DLT in the future $D_{y_d,n_d+m_d}$, which can be summarized in Table 2.7. Specifically, suppose a new patient is eligible for enrollment, the detailed rolling enrollment rules are described below.

I. If $m_d = 0$, i.e., all the patients enrolled at dose level $d$ have completed their followup with definitive outcomes, assign the new patient according to $D_{y_d,n_d}$, the decision of mTPI-2 when $y_d$ out of $n_d$ patients experience DLT outcomes.

II. If $0 < m_d \leq C$, i.e., some patients are still being followed without outcomes, consider three cases:

1. If $D_{y_d,n_d}$ is $D$, consider the following two cases:
   (a) if $D_{y_d,n_d+m_d}$ is $D$, de-escalate to dose level $(d - 1)$; apply the run-in enrollment if dose $(d - 1)$ is a new dose or re-apply the rolling enrollment if it has been used before;
   (b) else, the decision is $S$ and continue patient enrollment at dose $d$.

2. If $D_{y_d,n_d}$ is $S$, consider the following two cases:
   (a) if $D_{y_d,n_d+m_d}$ is $S$, assign the new patient to $d$;
   (b) if $D_{y_d,n_d+m_d}$ is $E$,
      i. if $k_d < 3$, enroll the next patient at dose $d$;
      ii. if $k_d \geq 3$, suspend the enrollment until more patients have observed their outcomes at dose $d$. Then recalculate the $m_d$ value and re-apply I or II.

3. If $D_{y_d,n_d}$ is $E$, consider the following two cases:
   (a) if $D_{y_d+m_d,n_d+m_d}$ is $E$, escalate to dose level $(d + 1)$; apply the run-in enrollment if dose $(d + 1)$ is a new dose or re-apply the rolling enrollment if it has been used before.
   (b) else,
      i. if $k_d < 3$, enroll the next patient to dose $d$;
      ii. if $k_d \geq 3$, suspend the enrollment until more patients have observed their outcomes at dose $d$. Then recalculate the $m_d$ value and re-apply I or II.

III. If $m_d > C$, suspend the enrollment until more patients have observed outcomes at dose $d$. 

Your trial designs anywhere, anytime.
2.3.5.3 Safety Rules

For trial safety, two additional rules are applied.

– [Rule 1: Dose Exclusion] If the current dose is considered excessively toxic, i.e., \( \text{Prob}\{p_d > p_T \mid Data\} > \xi \), where the threshold \( \xi \) is close to 1, say 0.95, the current and all higher doses are excluded and never be used again in the remainder of the trial.

– [Rule 2: Early Stop] If the current dose is the lowest dose (first dose) and is considered excessively toxic according to Rule 1, early stop the trial for safety.

Here, \( \text{Prob}\{p_d > p_T \mid Data\} \) is a function of the cumulative beta distribution \( \text{Beta}(\alpha_0 + y_d, \beta_0 + n_d - y_d) \). In U-Design, we use \( \alpha_0 = \beta_0 = 1 \). One thing needs to be noticed, at the time a dose is deemed unsafe and suspended, there may be some patients with pending outcomes at this dose level. Once these data are observed later, if the safety rule is no longer violated, this dose could be reopened again for further evaluation.

2.3.5.4 Trial Termination

The R-TPI design continues until any of the following conditions is satisfied:

1. If the prespecified maximum total sample size is reached;
2. If the lowest dose shows excessive toxicity according to Rule 2; In this case, the trial is early stopped and the MTD cannot be determined;
3. Optional: ad-hoc rules of maximum number of patients at a dose, denoted by \( K \) (\( K < n \)):
   - If the mTPI-2 decision is “S”, to stay at the current dose, and the current dose has enrolled \( K \) patients;
   - If the mTPI-2 decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled \( K \) patients;
   - If the mTPI-2 decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled \( K \) patients.

2.3.5.5 MTD Selection

Once all the enrolled patients complete the DLT observation and the trial is not stopped early, the R-TPI design applies an isotonic regression to select the MTD. Follow the steps as below.

1. Compute the isotonically transformed posterior means of DLT probabilities for all the dose levels in the following two steps.
   (a) Using the accumulated safety information about \( y_d \) and \( n_d \) for \( d = 1, \ldots, D \), compute
the posterior mean and variance for all the dose levels, \( \{ \tilde{p}_1, \cdots, \tilde{p}_D \} \) and \( \{ v_1, \cdots, v_D \} \).

Here in U-Design, an independent prior \( \text{beta}(0.005, 0.005) \) is used to compute the posterior mean and variance.

(b) Compute isotonic regression estimates of the posterior mean by solving the optimization problem, minimizing \( \sum_{d=1}^{D} (\hat{p}_d - \tilde{p}_d)^2 / v_d \) subject to \( \hat{p}_j \geq \hat{p}_k \), for \( j > k \). Such optimization can be done using the pooled adjacent violators algorithm (PAVA), the estimated DLT probabilities satisfying the order constraint is obtained, denoted by \( \{ \hat{p}_1, \cdots, \hat{p}_D \} \).

2. Among all the tried doses for which \( \text{Prob}\{ p_d > p_T \mid \text{Data} \} < \xi \) and \( \hat{p}_d \leq p_T + \epsilon_2 \), select as the estimated MTD the dose with the smallest difference \( |\hat{p}_d - p_T| \). That is, the estimated MTD is \( d^* = \arg\max_d |\hat{p}_d - p_T| \).

3. In case of a tie (i.e., two or more doses have the smallest difference),

(a) If there is at least one dose lower the target \( p_T \) among all the tied doses, choose the highest dose among those as the estimated MTD;

(b) Otherwise, choose the lowest dose among the tied doses as the estimated MTD.

2.3.5.6 R-TPI Decision Table

The R-TPI design requires users to provide the value of the target toxicity rate \( p_T \) and two small fractions, \( \epsilon_1 \) and \( \epsilon_2 \). The \( p_T \) value can be easily elicited from the trial clinician. The values of \( \epsilon_1 \) and \( \epsilon_2 \) are can be set at 0.05 as the default (Ji et al., 2010) or elicited by asking the clinician the lower and higher bound of the DLT rate that would still be considered as close to \( p_T \). Also we need to elicit the value of \( C \) to control the speed of patient accrual. With the provided values of \( p_T, \epsilon_1, \epsilon_2, \) and \( C \), one can generate the R-TPI decision table prior to the trial. Therefore, even though a model-based rolling design, R-TPI exhibits the same simplicity and transparency as rule-based methods.

We provide the decision table of up to seven patients for R-TPI with target DLT rate \( p_T \) equal to 0.3, \( \epsilon_1 = \epsilon_2 = 0.05 \), and \( C = 3 \), as an example. See Table 2.8.
2.3. Statistical Methods Review

2.3.5. The Rolling Toxicity Probability Interval (R-TPI) Design

Table 2.8: R-PTI Decision Table with $p_T = 0.3$, $\epsilon_1 = \epsilon_2 = 0.05$, and $C = 3$.

<table>
<thead>
<tr>
<th>Observed data at dose $d$</th>
<th>R-TPI Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_d + m_d$</td>
<td>$y_d$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>&gt; 0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>&gt; 1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>&gt; 1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0,1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>&gt; 1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0,1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>&gt; 2</td>
</tr>
</tbody>
</table>
3. **Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment**

3.1 **Introduction**

Gene therapies and adoptive cell therapies (ACTs), such as the chimeric antigen receptor (CAR) T-cell therapy, have demonstrated promising therapeutic effects in oncology patients. An important and distinct feature of some ACTs is that the probability of response may not increase with dose, which is normally seen for cytotoxic cancer therapeutics. For example, Porter et al. (2011) has shown that increased dose of CAR T-cells does not necessarily lead to higher efficacy. Because of the potential non-monotone relationship between response and dose, traditional phase 1 dose-finding designs searching for the maximum tolerated dose (MTD), like i3+3 (Liu et al., 2020) and mTPI-2 (Guo et al., 2017) designs, are not suitable to ACTs. For example, the best efficacious dose may be lower than the MTD as higher doses may not lead to higher efficacy.

To this end, the U-Design introduces the **Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment** module which consists of five novel statistical designs for gene and cell therapeutics dose-finding trials. The module performs trial simulations allowing head-to-head comparison of multiple designs, so that users may select the best design for their own clinical trials. The included novel designs are Ji3+3 (Lin and Ji, 2020b), PRINTE (Lin and Ji, 2020a), TEPI (Li et al., 2017), EffTox (Thall and Cook, 2004) and UBOIN (Zhou et al., 2019), all of which use joint toxicity and efficacy outcomes as endpoints for dose finding. The goal is to identify the optimal biological dose (OBD) that possesses high efficacy and safety simultaneously. As with all other U-Design modules involving trial simulation, below we provide detailed guidance...
3.1. Introduction

on setting up simulation for design comparison, and visualising simulation results (operating characteristics). In addition, the decision tables generation and the OBD selection are incorporated in this module so that users may generate the decision tables to guide trial conduct and estimate the OBD after trial completion. All the details are provided next.
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

3.2 User Interface and Tutorial

3.2.1 Overview

Entering the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment page, users will see four main tabs: Simulation Setup, Simulation Results, Decision table, and OBD selection. These four tabs allow users to conduct simulations and visualize/download simulation results, generate decision tables for trial conduct, and select OBD after trial is completed.

In the Simulation Setup tab, there are three steps (Figure 3.1): 1) Set trial parameters, 2) Select designs, and 3) Generate scenarios. Users need to complete the current step to get access to the next one. Upon completing steps 1-3, users click the “Launch Simulation” button at the bottom of the page to submit the simulations to the cloud for computation. Users may also click the “Reset” button next to Launch Simulation to clear all settings. After the simulation is launched, the results of simulations will be displayed in the Simulation Results tab. The simulation process can be monitored in real time at the top of the Simulation Results tab. Detailed steps of using this module are elaborated next in \(3.2.2-3.2.5\).
3.2. User Interface and Tutorial

3.2.1. Overview

Figure 3.1: Simulation Setup in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.
Module 3. Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment

3.2.2 Simulation Setup

In the Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment module, U-Design provides five designs, Ji3+3, PRINTE, TEPI, Efftox, UBOIN, for simulation. Users can choose up to four designs for head-to-head comparison in the Simulation Setup tab each time. Three steps of simulation set up are needed.

3.2.2.1 Step 1: Set trial parameters

Specify the number of simulations ($n_{sim}$) and the random seed of simulation ($R_{seed}$). Specify the target toxicity probability ($p_T$) and minimum acceptable efficacy ($q_E$) for the simulated trials and select a number of doses ($n_{dose}$) from the dropdown box. Click the “Apply” button to apply the settings. See Figure 3.2. Hover mouse over each trial parameter, and a description will be displayed explaining the meaning of the parameter. The detailed explanation on U-Design interface of the above four input arguments is provided in Table 3.1.

![Figure 3.2: Set trial parameters in the Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment module.](image)

Upon clicking the “Apply” button, a table of actual dosage will be displayed. Specify the dosage of each dose level in the table. (Figure 3.3) This is only needed if the EffTox design is selected in Step 2 next. If EffTox is not going to be selected, leave the table unchanged and move to Step 2.

3.2.2.2 Step 2: Select designs

To select a design, click the button with the design’s name on it. Up to four designs may be selected for head-to-head comparison.

Check the “Apply Stopping Rule” box to apply an ad-hoc stopping rule that stops the trial if a maximum number of patients has been enrolled at a single dose. See the detailed rules in §3.3.
3.2. User Interface and Tutorial

3.2.2. Simulation Setup

Figure 3.3: Selecting actual dosage in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

Click the “Draw” button to plot a contour map of the utility function. The horizontal axis represents efficacy and the vertical axis represents toxicity. See Figure 3.4.

Click the “Apply” button of all the designs before launching simulations to apply all settings.

Click the “Delete” button to remove the selected designs.

Design parameters can be modified in the input box. Hover mouse over each design parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 3.2.
Table 3.1: Input parameters for trials parameters in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{sim}$</td>
<td>The number of simulated trials</td>
<td>The maximum number of simulated trials allowed is 10,000. Default value is 1,000.</td>
</tr>
<tr>
<td>$R_{seed}$</td>
<td>The random seed of simulation</td>
<td>A random seed is a number used to initialize a pseudorandom number generator in the simulation. Default value is 32432.</td>
</tr>
<tr>
<td>$p_T$</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). Default value is 0.3.</td>
</tr>
<tr>
<td>$q_E$</td>
<td>Minimum acceptable efficacy</td>
<td>The minimum acceptable efficacy used in the futility rule. A dose is considered not promising if the efficacy rate is unlikely to be larger than $q_E$. Default value is 0.2.</td>
</tr>
<tr>
<td>$n_{dose}$</td>
<td>Number of doses</td>
<td>The number of doses in the trial.</td>
</tr>
</tbody>
</table>
3.2. User Interface and Tutorial

3.2.2. Simulation Setup

Figure 3.4: Select designs in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.
**Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment**

**Table 3.2:** Input parameters for designs in the **Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment** module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_{\text{start}}) (all designs)</td>
<td>Starting dose level</td>
<td>The starting dose level in the simulated trials. Default value is 1.</td>
</tr>
<tr>
<td>(n) (all designs)</td>
<td>Sample size</td>
<td>The maximum number of patients to be treated in the trial. The upper limit is set at 100 since the number of patients that are enrolled in phase I clinical trial is typically small. Default value is 30.</td>
</tr>
<tr>
<td>(n_{\text{cohort}}) (all designs)</td>
<td>Cohort size</td>
<td>The number of patients in each cohort. Default value is 3.</td>
</tr>
<tr>
<td>(K) (all designs)</td>
<td>Maximum number of patients at a dose level</td>
<td>A number used in the “Stopping Rule” that stops a trial if 1) the dose-assignment decision is to escalate to the next higher dose and there has been (K) patients enrolled at that dose; or 2) the dose-assignment decision is to stay at the current dose and there has been (K) patients enrolled at that dose; or 3) if the dose-assignment decision is to de-escalate to the previous lower dose and there has been (K) patients enrolled at that dose; Default value is 12.</td>
</tr>
<tr>
<td>(p_{\text{cut}}) (all designs)</td>
<td>Cutoff probability for futility rule</td>
<td>A cutoff probability used in the safety rule. Exclude dose (d) if (Pr(p_d &lt; p_T</td>
</tr>
<tr>
<td>(q_{\text{cut}}) (all designs)</td>
<td>Cutoff probability for efficacy rule</td>
<td>A cutoff probability used in the futility rule. Exclude dose (d) if (Pr(q_d &lt; q_E</td>
</tr>
<tr>
<td>(p_E) (Ji3+3, PRINTE)</td>
<td>Target efficacy probability</td>
<td>The lower bound of the response probability for the treatment to be considered promising and warrant further clinical development. Default value is 0.4.</td>
</tr>
<tr>
<td>(\epsilon_1, \epsilon_2) (Ji3+3, PRINTE)</td>
<td>(\epsilon_1): lower margin (\epsilon_2): higher margin</td>
<td>Two small fractions used to define the equivalence interval of the MTD. Any dose with a toxicity probability falling into the interval ([p_T - \epsilon_1, p_T + \epsilon_2]) is considered an acceptable dose MTD. Default values for both are 0.05.</td>
</tr>
</tbody>
</table>
### 3.2. User Interface and Tutorial
#### 3.2.2. Simulation Setup

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1^<em>, p_2^</em>$</td>
<td>Prespecified cutoff values in utility function on toxicity. Cutoff values in utility function for toxicity. The toxicity utility score is 1 when $p &lt; p_1^<em>$, is 0 when $p &gt; p_2^</em>$ and linearly decreases when $p$ is between $(p_1^<em>, p_2^</em>)$. Default values are 0.2 and 0.4.</td>
</tr>
<tr>
<td>$q_1^<em>, q_2^</em>$</td>
<td>Prespecified cutoff values in utility function on efficacy. Cutoff values in utility function for efficacy. The efficacy utility score is 0 when $p &lt; p_1^<em>$, is 1 when $p &gt; p_2^</em>$ and linearly increases when $p$ is between $(p_1^<em>, p_2^</em>)$. Default values are 0.2 and 0.6.</td>
</tr>
<tr>
<td>$p_{grad}$</td>
<td>Cutoff probability for a dose to be considered as OBD. A cutoff value used when choosing OBD. If the posterior probability of utility function lying in the admissible utility region is below $p_{grad}$, no OBD will be selected and the trial ends without selecting an optimal dose. Default value is 0.2.</td>
</tr>
<tr>
<td>$a_1, b_1$</td>
<td>Prior beta distribution parameters of toxicity rate. The parameters in the prior beta distribution of toxicity rate, $Beta(a_1, b_1)$. Default values for both are 1 to be conservative, since a $Beta(1,1)$ prior implies a prior a dose has a toxicity rate of 0.5 with effective sample size of 0.5.</td>
</tr>
<tr>
<td>$a_2, b_2$</td>
<td>Prior beta distribution parameters of efficacy rate. The parameters in the prior beta distribution of efficacy rate, $Beta(a_2, b_2)$. Default values for both are 0.5, which is Jeffreys’s prior (Jeffreys, 1946).</td>
</tr>
<tr>
<td>$s_1$</td>
<td>Maximum sample size in one dose at stage 1. The maximum number of patients to be treated in one dose at stage 1. Move to stage 2 when the number of patients treated on one of the doses reaches $s_1$. A value between 9 and 15 generally yields good operating characteristics. Default value is 12.</td>
</tr>
<tr>
<td>$s_2$</td>
<td>Maximum sample size at one dose at stage 2. The maximum number of patients to be treated in one dose at stage 2. Stop the trial and choose OBD when the number of patients treated at one of the doses reaches $s_2$. For most trials, a value between 18 and 24 is a reasonable choice for $s_2$. Default value is 18.</td>
</tr>
</tbody>
</table>
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

<table>
<thead>
<tr>
<th>Pick The Winner, Adaptive Randomization (UBOIN)</th>
<th>Methods to select next dose</th>
<th>Pick The Winner: The pick-the-winner (PW) approach deterministically assigning the next cohort of patients to a dose that has the largest posterior mean utility. Adaptive Randomization: The adaptive randomization (AR) approach adaptively randomizes the next cohort of patients to a dose with probability proportional to its posterior mean utility.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi^<em>_1, E, \pi^</em>_2, T, \pi^<em>_3, E, \pi^</em>_3, T) (EffTox)</td>
<td>Parameters in the desirable trade-off target values</td>
<td>(\pi^<em>_1, E) is the smallest efficacy probability that the physician would consider desirable if toxicity were impossible. (\pi^</em>_2, T) is the maximum desirable value of toxicity if the efficacy were 1. Set (\pi^<em>_1, E, \pi^</em>_2, T, \pi^<em>_3, E, \pi^</em>_3, T) so that (\pi^<em>_1 = (\pi^</em>_1, E, 0), \pi^<em>_2 = (1, \pi^</em>_2, T), \pi^<em>_3 = (\pi^</em>_3, E, \pi^*_3, T)) Default values are 0.15, 0.6, 0.25, 0.3.</td>
</tr>
</tbody>
</table>

Your trial designs anywhere, anytime.
3.2. User Interface and Tutorial

3.2.2. Simulation Setup

3.2.2.3 Step 3: Generate scenarios

There are two ways to generate scenarios, automatically (in the “Auto Generation” tab) or through manual construction (in the “Manual Construction” tab). Users could also manually add or delete scenarios. Once scenarios are generated, click the button “Submit” to notify the software that the scenarios are final, then click the “Launch Simulation” button at the bottom of the page to run $n_{\text{sim}}$ (set in step 1) simulations, for each scenario and selected design (set in step 2), using the $p_T$ and $q_E$ values. (set in step 1).

**Auto Generation**  Click the “Auto Generation” button and six diverse scenarios will be created automatically, each of which contains the true toxicity probabilities for $n_{\text{dose}}$ dose levels. These generated scenarios are displayed (Figure 3.5). One can click the button to delete any scenario.

**Manual Construction**  A list of toxicity/efficacy probabilities are displayed. Click “Add” to add an empty, editable row of toxicity or efficacy probabilities. Click the button to delete the row. Click “Delete All” to delete all the rows.

Check the “Select” box in the front to select the row of toxicity or efficacy probabilities. Click “Select All” to select all the toxicity or efficacy rows.

Upon selection, click “Generate” to generate scenarios which will combine existing rows of toxicity and efficacy probabilities. The scenarios will be displayed in. (Figure 3.7)

Once the scenarios are generated, clicking the button will delete a scenario. Clicking “Delete All” will delete all the scenarios. Click the “Submit” button to notify the software that all the scenarios are final (Figure 3.7). If there are duplicated scenarios in the list, a message will be displayed on the website to indicate that the duplicated scenarios have been removed. Click the “OK” button to proceed to launch simulation. (Figure 3.8)

3.2.2.4 Launch Simulation

Once the above Steps 1-3 are completed, users can conduct simulated clinical trials to examine the operating characteristics of the selected designs using the selected scenarios, by clicking the “Launch Simulation” button at the bottom of Simulation Setup tab (Figures 3.7). A green “Launch Successful” message will be displayed on the website (Figure 3.9) to indicate that the simulation has been successfully launched. Users may click the “Proceed to Simulation Results” button in the pop-up box to track the simulation processing status and simulation results.
Figure 3.5: Automatically generated scenarios in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.
Figure 3.6: Selecting toxicity and efficacy in the **Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment** module.
Module 3. Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment

Figure 3.7: Selecting scenarios in the Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment module.

Figure 3.8: Removing the duplicated scenarios in the Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment module.
Figure 3.9: “Launch Successful” message after launching simulation in the Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment module.
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

3.2.3 Simulation Results

In the Simulation Results tab, users can view and delete the simulation progress and simulation results (§3.2.3.1), restore the simulation settings if needed (§3.2.3.2), and download intelligent simulation reports (§3.2.3.3). Specifically, all the simulation results (figures and tables) can be downloaded in Word format, accompanying the statistical sections in a trial protocol. Hereinafter, we use simulation results and operating characteristics interchangeably.

3.2.3.1 View simulation results

In the Simulation Results tab, the Running Simulations panel exhibits the progress of ongoing simulation (Figure 3.10). The ongoing simulations are displayed in ascending order by the launch time. Click the icon “×” to delete the corresponding simulation.

![Running Simulations](image)

**Figure 3.10:** Simulation progress in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

Once the simulations are completed, the Running Simulations panel in Figure 3.10 will disappear, green “simulation result created” massages will appear instead and stay at the same place of the Running Simulations panel unless explicitly dismissed by clicking the icon “×” at the end of the corresponding row, and the simulation results will be automatically loaded into the Simulation History panel (Figure 3.11), with the blue bold Unread term shown to indicate new results. All the previously completed simulations are also listed in the Simulation History panel. Simulation results for other modules can also be viewed under the Simulation History by dropping down the “Select a module” button (Figure 3.11).
3.2. User Interface and Tutorial
3.2.3. Simulation Results

Figure 3.11: Simulation Results in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

Click the button to unfold the simulation results (Figure 3.12). The design settings are firstly displayed at the top of each simulation study (Figure 3.12). Then the results of simulation are shown as plots and tables below. And one can also click the button to delete the selected simulation results.

Figure 3.12: View the simulation results in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

Simulation Result Tables
Full simulation results are presented in tabular format arranged by scenarios (Figure 3.13).

In the upper part of Figure 3.13, the first three columns summarize dose levels, their true toxicity and true efficacy probabilities; the remaining columns report four dose-specific summary statistics from the simulations: selection probability, average number of patients treated, average number of toxicities (i.e., DLTs), along with their standard deviations, and average number of responses,
along with their standard deviations, at each dose level. Specifically, they are

1) **Selection Prob.** The proportion of simulated trials that select each dose level as the MTD.
2) **Average # of Patients Treated (s.d.).** The average number of patients treated at each dose level and its standard deviation.
3) **Average # of Toxicities (s.d.).** The average number of patients experienced DLT at each dose level and its standard deviation.
4) **Average # of Responses (s.d.).** The average number of patients observed efficacy response at each dose level and its standard deviation.

The true OBD(s) of the scenario is (are) highlighted by the orange bar. The true OBD is defined as the dose that achieves the highest utility, which could be calculated using true toxicity, efficacy probabilities and the utility function.

In the lower part of Figure 3.13, more trial-specific summary statistics are reported, mainly from five aspects: **OBD Selection, Subjects Assignment**, **Trial Toxicity**, **Trial Stopping** and **Trial Sample Size.** Specifically, they are

- **OBD Selection**
  - **Prob. of Selecting OBD.** The proportion of simulated trials that select the true OBD at the end of the trial. The higher the value, the better the design.
  - **Prob. of Selecting Does-over-MTD.** The proportion of simulated trials that select the doses higher than the true MTD at the end of the trial. The lower the value, the better the safety of the design.
  - **Prob. of No Selection.** The proportion of simulated trials in which none of the dose levels are selected as the OBD. If a scenario does not have any OBD, this values is treated as the probability of selecting the true OBD.

- **Subjects Allocation**
  - **Prob. of Correct Allocation (s.d.).** The average proportion of patients who are correctly assigned to the true OBD by the design across all the simulated trials and its standard deviation. The higher the value, the better the design.
  - **Prob. of Overdosing Allocation (s.d.).** The average proportion of patients who are assigned to doses higher than the MTD by the design across all the simulated trials and its standard deviation. The lower the number, the better the safety of the design.

- **Trial Toxicity**
  - **Prob. of Toxicity.** The proportion of patients experiencing DLT across all the simulated trials. The lower the number, the fewer patients having DLTs under the design.
3.2. User Interface and Tutorial
3.2.3. Simulation Results

- **Trial Stopping**
  - **Prob. of Early Stopping Trial due to No admissible dose**: The proportion of simulated trials in which the trial is stopped because there is no admissible dose left. This means that all the doses have unacceptable toxicity or efficacy and are excluded by safety rule or futility rule.
  - **Prob. of Early Stopping Trial due to Reaching $K$**: The proportion of simulated trials in which the trial is stopped because the dose-assignment decision is to escalate/stay/de-escalate to a dose level but that dose has enrolled at least $K$ patients ($K < n$, e.g., $K = 12$).
  - **Prob. of Stopping Trial due to Reaching $n$**: The proportion of simulated trials in which the trial is stopped because the total number of patients enrolled and treated in a trial has reached or exceeded the pre-specified maximum sample size $n$.

- **Trial Sample Size**
  - **Average # of Patients Treated (s.d.)**: The average number of patients treated in the simulated trials and its standard deviation. Due to early stopping, this number is lower than or equal to $n$.

- **Statistics of UBOIN**
  - **Prob. of Entering Stage II**: The proportion of simulated trials in which the trial enters Stage II because the number of patients at one dose has reached or exceeded the pre-specified maximum sample size $s_1$ in Stage I.
  - **Average # of Patients Treated in Stage I**: The average number of patients treated in Stage I in the simulated trials.
  - **Average # of Patients Treated in Stage II**: The average number of patients treated in Stage II in the simulated trials.

When calculating the standard deviation, we use $n_{sim}$ as the denominator instead of $(n_{sim} - 1)$ in U-Design.
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

**Figure 3.13:** Simulation result tables in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>( \text{Scenario 1} )</th>
<th>( \text{Scenario 2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
<tr>
<td>( \text{Scenario 1} )</td>
<td>( \text{Scenario 2} )</td>
<td>( \text{Scenario 3} )</td>
</tr>
</tbody>
</table>

* The cell with orange background color indicates the TRUE CEB for each design.
3.2.3.2 Restore simulation setup

Users can restore the simulation settings from the simulation results by clicking the button at the upper right corner of each simulation results panel (yellow arrow in Figure 3.14). Upon clicking, the display will switch to the Simulation Setup page with the same simulation settings restored. This is useful to restore the old simulation settings for Reproducible results.

![Figure 3.14: Restore simulation setup and download simulation results in the Single-Agent Dose-Finding Designs with Efficacy&Toxicity Endpoints and Cohort Enrollment module.](image)

3.2.3.3 Download simulation results

There is a button at the upper right corner of each simulation results panel (green arrow in Figure 3.14). Click it to download a word file, which includes four parts:

- Part A: Complete simulation results under the designs and scenarios users added in the Simulation Setup tab;
- Part B:
- Part C: Detailed technical descriptions of the designs users added in the Simulation Setup tab;
- Part D: Reference

Users may select the required parts and modify them tailored for their trials or contact us via email (admin@laiyaconsulting.com) for consulting services.
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

3.2.4 Decision Tables

This function generates decision tables based on the Ji3+3, PRINTE, and TEPI designs, which can be used to conduct a dose-finding trial. Users can click the tabs to switch between the tables for the Ji3+3, PRINTE, and TEPI designs.

Manually type in the design settings for decision table generation (Figure 3.15). The parameters are the same as the ones in Step 2 (3.2.2.2) in the Simulation Setup tab. See detailed parameter descriptions in Table 3.2.

![Decision Table Generation](image)

**Figure 3.15:** Input parameters in the Generate Decision Table tab of Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

Click the “Generate” button to generate decision table (Figure 3.16). Decision tables are automatically generated for 3, 6, 9 and 12 patients at a dose in the panel below.

To generate a single decision table by specifying the number of patients treated at a dose $d$, set $n_d$ in the box and click the button “Add”. (Figure 3.16)
Figure 3.16: Decision tables in the Generate Decision Table tab of Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

For each decision table, the column represents the number of patients responses among those treated at the dose, and the row represents the number of patients who have experienced dose-limiting toxicity (DLT) events. Note that these are the counts of patients, not DLT events or responses. For example, column 3 and row 1 means that among the patients that have been treated at the current dose 3 of them experiences DLT, and 1 of them responses.

Each cell in the decision table provides the dose-assignment decision based on the readouts from the corresponding row and column. For example, for column 3 and row 1, i.e., 3 patients experience DLTs, and 1 patient has efficacy response, the decision is “EU”. The letters in the decision table represent different dose-assignment decisions as shown below:

- “E” stands for escalating to the next higher dose,
- “S∗” stands for staying at the current dose, or escalate to dose $d+1$ if $d$ is not the highest dose and $d+1$ is untried
- “S” stands for staying at the current dose.

Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

- “D” stands for de-escalating to the previous lower dose,
- “DUT” stands for de-escalating to the previous lower dose, and the current dose and its higher doses is deemed unacceptable due to severe toxicity and will not be used again in the study. If at the first dose level, users can choose to early-terminate the trial or not based on their own discretion.
- “EUE” stands for escalating to the higher dose and marking the current dose as unacceptable (due to futility) so that it will never be used again in the remainder of the trial.
- “DUE” stands for de-escalating to the previous lower dose, and the current dose is deemed unacceptable due to futility and will not be used again in the study.

Some additional detailed explanation of the decisions are provided in the decision table report. The meaning of the notations are shown below:

- The superscript * on DUE indicates that according to the Ji3+3 design, the decision is S and the current dose is deemed unacceptable due to futility. In this case, a decision S indicates a moderate or high toxicity probability, so the only sensible action is to de-escalate to the previous lower dose, and remove the current dose (due to futility) from the study.
- The superscript ** on DUT indicates that if the current dose is the first dose level, users can choose to early-terminate the trial or not based on their own discretion.

Click “DOWNLOAD ONE” to download a word file, which includes the design settings and the single decision table in the tab selected. Click “DOWNLOAD ALL” to download a word file, which includes the design settings and all the decision tables generated.
3.2.5 OBD Estimation

In this module, all designs aim to estimate the OBD when the trial is completed and the data is collected. The detailed statistical models for the included designs are described in §3.3.

First, select a design and provide corresponding model parameters. Second, select the number of doses \( n_{\text{dose}} \) from the dropdown box, and an editable table will be shown on the website (Figure 3.17). For the Ji3+3, PRINTE and TEPI design, provide the number of patients treated, the observed number of DLT events, and provide the observed number of efficacy events at each dose into the table; For the UBOIN and Efftox design, provide the observed number of patients who has no efficacy but DLT \((Y_E, Y_T) = (0, 1)\), no efficacy and no DLT \((Y_E, Y_T) = (0, 0)\), efficacy and DLT \((Y_E, Y_T) = (1, 1)\) and the number of patients who has no efficacy and no DLT \((Y_E, Y_T) = (1, 0)\). Click the “Generate” button to estimate the utilities of each dose and estimate the OBD for the trial. The estimated utility will be displayed in a table and the estimated OBD will be highlighted in green color as shown in Figure 3.18.

Hover mouse over each parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 3.2 in §3.2.2.2.
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

Figure 3.17: Input parameters in the OBD Estimation tab of Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.
3.2.5. OBD Estimation

* Figure 3.18: Determine the estimated OBD in the Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment module.

* The dose highlighted in green is selected as the OBD.
3.3 Statistical Methods Review

3.3.1 The Joint i3+3 (Ji3+3) Design

Ji3+3 is a rule-based phase I/II ACT dose-finding design proposed by Lin and Ji (2020b). Building upon i3+3 (Liu et al., 2020), Ji3+3 takes into account of both toxicity and efficacy outcomes in making dosing recommendations. Basically, the decision rules of the Ji3+3 design incorporate and extend the toxicity rules in i3+3 with a set of efficacy rules. Simulation results show that Ji3+3 outperforms existing designs when monotonic dose response assumption is violated, and achieves comparable performance when the assumption holds. Since Ji3+3 is a model-free design, it is transparent to physicians and simple to implement.

3.3.1.1 Dose-Finding Algorithm

Consider $D$ ascending doses in a single-agent ACT phase I trial. Due to ethical considerations, it is always assumed that the toxicity probability $p_d$ increases with dose level $d$, that is, $p_1 \leq \cdots \leq p_D$. However, the efficacy probability $q_d$ may increase initially and then reach a plateau from which minimal improvement or even decreasing efficacy may be seen with increasing dose. For this reason, we assume that $q_d$ is not monotone with $d$, and that $p_d$ and $q_d$ are independent. Suppose that dose $d$ is currently used in the trial and $n_d$ patients have already been allocated to dose $d$, with $x_d$ and $y_d$ patients experiencing toxicity and efficacy outcomes. Aggregating across all the doses, the trial data are denoted as $\text{Data} = \{(n_d, x_d, y_d), d = 1, \cdots, D\}$

Denote $p_T$ as the target toxicity rate, which is the probability of toxicity at the MTD; denote $p_E$ as the target efficacy rate. In Ji3+3, $[p_T - \epsilon_1, p_T + \epsilon_2]$ is defined as the Equivalence Interval (EI), where $(\epsilon_1, \epsilon_2)$ are two small fractions that account for the uncertainty around $p_T$. This allows doses whose toxicity probabilities differ from $p_T$ to be considered as the MTD. Given the observed data $\text{Data}$, the dose-finding algorithm of the Ji3+3 design is shown in Table 3.3. The algorithm follows these principles:

1. If there is lack of evidence for efficacy, escalate to achieve higher efficacy; else, stay at the current dose because it is considered to have sufficient efficacy.
2. For toxicity, the idea is to compare the observed toxicity rate $\frac{x_d}{n_d}$ with the EI.
   - If $\frac{x_d}{n_d}$ is below the EI, the dose is considered safe; if $\frac{x_d}{n_d}$ is inside the EI, the dose is considered to be close to the MTD; if $\frac{x_d}{n_d}$ is above the EI, the dose is considered not safe except when $\frac{x_{d-1}}{n_d}$ is below the EI.
3.3. Statistical Methods Review

3.3.1. The Joint i3+3 (Ji3+3) Design

Table 3.3: Schema of the Ji3+3 design.

<table>
<thead>
<tr>
<th>Eff cond.</th>
<th>Tox cond.</th>
<th>Next dose (Decision)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{x_d}{n_d} \leq p_E )</td>
<td>( \frac{x_d}{n_d} &lt; \text{EI} )</td>
<td>( d + 1 \ (E) )</td>
</tr>
<tr>
<td>( \frac{x_d}{n_d} \in \text{EI} )</td>
<td>( \frac{x_d}{n_d} &gt; \text{EI} ) &amp; ( \frac{x_{d-1}}{n_d} &lt; \text{EI} )</td>
<td>( d \ (S \text{ or } E^*) )</td>
</tr>
<tr>
<td>( \frac{x_d}{n_d} &gt; \text{EI} ) &amp; ( \frac{x_{d-1}}{n_d} \in \text{EI} )</td>
<td>( d - 1 \ (D) )</td>
<td></td>
</tr>
<tr>
<td>( \frac{x_d}{n_d} &gt; \text{EI} ) &amp; ( \frac{x_{d-1}}{n_d} &gt; \text{EI} )</td>
<td>( d - 1 \ (D) )</td>
<td></td>
</tr>
</tbody>
</table>

\( \frac{x_d}{n_d} > p_E \)

<table>
<thead>
<tr>
<th>Eff cond.</th>
<th>Tox cond.</th>
<th>Next dose (Decision)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{x_d}{n_d} &lt; \text{EI} )</td>
<td>( \frac{x_d}{n_d} \in \text{EI} )</td>
<td>( d \ (S) )</td>
</tr>
<tr>
<td>( \frac{x_d}{n_d} &gt; \text{EI} ) &amp; ( \frac{x_{d-1}}{n_d} &lt; \text{EI} )</td>
<td>( d \ (S) )</td>
<td></td>
</tr>
<tr>
<td>( \frac{x_d}{n_d} &gt; \text{EI} ) &amp; ( \frac{x_{d-1}}{n_d} \in \text{EI} )</td>
<td>( d - 1 \ (D) )</td>
<td></td>
</tr>
<tr>
<td>( \frac{x_d}{n_d} &gt; \text{EI} ) &amp; ( \frac{x_{d-1}}{n_d} &gt; \text{EI} )</td>
<td>( d - 1 \ (D) )</td>
<td></td>
</tr>
</tbody>
</table>

*: Escalate to dose \( d + 1 \) if \( n_{d+1} = 0 \).

– When \( \frac{x_{d-1}}{n_d} \) is below the EI and \( \frac{x_d}{n_d} \) is above the EI, the data is noisy since increment of one toxicity event renders the observed toxicity rate to jump from below the EI to above the EI. In other words, the observed data is not very informative because change of one toxicity event can greatly influence the toxicity estimate.

Consider an example. Suppose EI = \([0.2, 0.3]\) with \( x_d = 1 \) and \( n_d = 3 \). Even though \( \frac{x_d}{n_d} = \frac{1}{3} \) is above the EI, \( \frac{x_{d-1}}{n_d} = \frac{0}{3} \) is below the EI. And therefore, dose \( d \) should not be considered as above the MTD.

3. Intersecting the two dosing principles for toxicity and efficacy, and taking the more conservative decision between the two, we arrive at the decisions in Table 3.3.

4. When \( d \) is the highest dose or lowest dose, the above rules are modified as special cases,

– If the current dose is the highest dose, decision E (escalate and treat the next cohort of patients at the next higher dose) should be replaced with decision S (stay and continue to enroll patients at the current dose), since there is no dose to escalate to.

– Similarly, if the current dose is the lowest dose, decision D (de-escalate to the next lower dose) should be replaced with S since there is no dose to de-escalate to.
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

Safety and futility rules

- Safety rule: if \( \Pr(p_d > p_T \mid x_d, n_d) > p_{\text{cut}} \) for a \( p_{\text{cut}} \) close to 1 (say, 0.95), exclude doses \( d, d+1, \cdots, D \), from future use in the trial; treat the next cohort of patients at dose \((d-1)\).

- Futility rule: if \( \Pr(q_d < q_E \mid y_d, n_d) > q_{\text{cut}} \) for a \( q_{\text{cut}} \) close to 1 (say, 0.7), where \( q_E \) is the minimum acceptable probability of efficacy, then exclude dose \( d \) from future use in the trial.

Here, \( q_E \) is the reference efficacy rate, e.g., the efficacy rate of standard care.

Note that, here we assume the prior for each \( p_d \) follows an independent \( \text{beta}(a_1, b_1) \), and the prior for each \( q_d \) follows an independent \( \text{beta}(a_2, b_2) \), where \( \text{beta}(\alpha, \beta) \) denotes a beta distribution with mean \( \alpha/(\alpha + \beta) \). The posterior distributions for \( p_d \) and \( q_d \) in the above rules are \( \text{beta}(a_1 + x_d, b_1 + n_d - x_d) \) and \( \text{beta}(a_2 + y_d, b_2 + n_d - y_d) \), respectively.

Stopping rules

The trial is stopped if

1. the prespecified maximum total sample size \( n \) is reached; or
2. the lowest dose shows excessive toxicity according to the safety rule; In this case, the trial is early stopped and the MTD cannot be determined; or
3. optional:
   - the Ji3+3 decision is “S”, to stay at the current dose, and the current dose has enrolled \( K \) patients;
   - the Ji3+3 decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled \( K \) patients;
   - the Ji3+3 decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled \( K \) patients.

3.3.1.2 Dose Selection

At the end of the trial, Ji3+3 chooses the OBD using a joint utility score \( U(p, q) = f_1(p)f_2(q) \) (suppressing dose \( d \) in the notation), which takes the product of toxicity utility \( f_1(p) \) in (3.1) and efficacy utility \( f_2(q) \) in (3.2).

\[
f_1(p) = \begin{cases} 
1, & p \in (0, p_1^*), \\
1 - \frac{p - p_1^*}{p_2^* - p_1^*}, & p \in (p_1^*, p_2^*), \\
0, & p \in (p_2^*, 1). 
\end{cases} \tag{3.1}
\]
3.3. Statistical Methods Review

3.3.1. The Joint i3+3 (Ji3+3) Design

For toxicity, define two thresholds \( p_1^* \) and \( p_2^* \) such that the toxicity utility score is 1 when \( p < p_1^* \), 0 when \( p > p_2^* \), and linearly decreases when \( p \) is between \( (p_1^*, p_2^*) \). For efficacy, define two thresholds \( q_1^* \) and \( q_2^* \) such that the efficacy utility score is 0 when \( q < q_1^* \), is 1 when \( q > q_2^* \), and linearly increases when \( q \) is between \( (q_1^*, q_2^*) \). The OBD is selected according to the following process.

1. We generate a total of \( T \) random samples, \( \{ p_d^{(t)}, t = 1, \cdots , T \} \) and \( \{ q_d^{(t)}, t = 1, \cdots , T \} \), from the posterior distributions \( \text{beta}(a_0 + x_d, b_0 + n_d - x_d) \) and \( \text{beta}(a_0 + y_d, b_0 + n_d - y_d) \) for each dose \( d \), respectively. Here, U-Design sets \( a_0 = b_0 = 0.005 \) and \( T = 1000. \)

2. For toxicity probabilities of all doses in each sample \( t \), \( p^{(t)} = (p_1^{(t)}, \cdots , p_D^{(t)}) \), we perform isotonic transformation using the pool adjacent violators algorithm (PAVA; Mair et al. 2009) on \( p^{(t)} \) to obtain \( \tilde{p}^{(t)} = (\tilde{p}_1^{(t)}, \cdots , \tilde{p}_d^{(t)}) \), where \( \tilde{p}_i^{(t)} \leq \tilde{p}_j^{(t)} \) if \( i < j \).

3. We propose a probabilistic inference for selecting the OBD and avoid selecting doses with low utility. Define an admissible probability region (APR) \( A(p, q) = \{ (p, q) \mid p \in (0, p_T], q \in [q_E, 1) \} \). Then the OBD is selected only from the candidate dose set \( \mathcal{A} \)

\[
\mathcal{A} = \{ d \mid p_{in,d} \geq p_{\text{grad}}, n_d > 0, d = 1, \cdots , D \},
\]

where \( p_{in,d} = \Pr \{ (p_d, q_d) \in \text{APR} \mid \text{Data} \} \) is the posterior probability that dose \( d \) belongs to APR and \( p_{\text{grad}} \) is a small value (say, 0.1). We use a simple a simple numerical approximation approach to compute \( p_{in,d} \) given by

\[
\hat{p}_{in,d} = \frac{1}{T} \sum_{t=1}^{T} 1 \{ (\tilde{p}_d^{(t)}, q_d^{(t)}) \in \text{APR} \}.
\]

4. The final selected dose \( d^* \) is the one that maximizes the utility score \( U(p_d, q_d) \). That is,

\[
d^* = \arg\max_{d \in \mathcal{A}} \hat{E}[U(p_d, q_d) \mid \text{Data}],
\]

where

\[
\hat{E}[U(p_d, q_d) \mid \text{Data}] = \frac{1}{T} \sum_{t=1}^{T} U(\tilde{p}_d^{(t)}, q_d^{(t)}).
\]
3.3.2 The Toxicity and Efficacy Probability Interval (TEPI) Design

TEPI, proposed in Li et al. (2017), is a practical dose-finding design for ACT trials that incorporates both toxicity and efficacy data. It is a natural extension of mTPI by adding the efficacy interval into the dose-finding model. TEPI partitions the unit intervals \((0, 1)\) for both the toxicity probability \(p_i\) and efficacy probability \(q_i\) into subintervals, denoted as \((a, b)\) and \((c, d)\), respectively. Then it uses beta-binomial models to estimate the efficacy and toxicity probability and makes dosing-decisions based on the joint unit probability mass (JUPM) of the interval combinations \((a, b) \times (c, d)\). TEPI is transparent to clinicians and simple to implement in practice.

3.3.2.1 Elicited decision table

The dose-finding algorithm of TEPI is based on a clinician-elicited decision table in terms of efficacy and toxicity probability intervals. The procedures of eliciting the decision table are as follows.

Consider \(D\) ascending doses in a single-agent ACT phase I trial. Due to ethical considerations, it is always assumed that the toxicity probability \(p_d\) increases with dose level \(d\), that is, \(p_1 \leq \cdots \leq p_D\). However, the efficacy probability \(q_d\) may increase initially and then reach a plateau from which minimal improvement or even decreasing efficacy may be seen with increasing dose. For this reason, we assume that \(q_d\) is not monotone with \(d\), and that \(p_d\) and \(q_d\) are independent. Suppose that dose \(d\) is currently used in the trial and \(n_d\) patients have already been allocated to dose \(d\), with \(x_d\) and \(y_d\) patients experiencing toxicity and efficacy outcomes. Aggregating across all the doses, the trial data are denoted as \(Data = \{(n_d, x_d, y_d), d = 1, \cdots, D\}\)

Partition the unit intervals \((0, 1)\) for \(p_d\) and \(q_d\) into four subintervals. Denoting \((a, b)\) and \((c, d)\) a subinterval in the partition for \(p_d\) and \(q_d\) respectively, where

\[
(a, b) \in \left\{ (0, t_1), (t_1, t_2), (t_2, t_3), (t_3, 1) \right\},
\]

\[
(c, d) \in \left\{ (0, e_1), (e_1, e_2), (e_2, e_3), (e_3, 1) \right\}.
\]

The interval combinations \((a, b) \times (c, d)\) form the basis for dose-finding decisions, with each combination corresponding to a specific decision, such as dose escalation or de-escalation. U-Design uses a default fixed decision for each interval combination, see Table 3.4.

In order to formulate this table, it is required to determine: (i) bounds of efficacy rate interval, \(e_1, e_2, e_3\), and (ii) bounds of toxicity rate interval, \(t_1, t_2, t_3\).
3.3. Statistical Methods Review

3.3.2. The Toxicity and Efficacy Probability Interval (TEPI) Design

Table 3.4: An default decision table for each interval combination.

<table>
<thead>
<tr>
<th>Toxicity Rate</th>
<th>Efficacy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>(0, t₁)</td>
</tr>
<tr>
<td>Moderate</td>
<td>(t₁, t₂)</td>
</tr>
<tr>
<td>High</td>
<td>(t₂, t₃)</td>
</tr>
<tr>
<td>Unacceptable</td>
<td>(t₃, 1)</td>
</tr>
</tbody>
</table>

Note: "E", "S" and "D" denote escalation, stay and de-escalation, respectively.

3.3.2.2 Dose-finding Algorithm

Building upon the preset table, we set up a local decision-theoretic framework and derive a Bayes rule. Here, local means that the framework focuses on the optimal decision to be made for the current dose instead of the trial. We show that the Bayes rule is equivalent to computing the joint unit probability mass (JUPM) for the toxicity and efficacy probability intervals. For a given region A, the JUPM is defined as the ratio between the probability of the region and the size of the region. Considering the two-dimensional unit square \((0, 1) \times (0, 1)\) in the real space, the JUPM for each interval combination \((a, b) \times (c, d)\) is

\[
\text{JUPM}^{(c,d)}_{(a,b)} = \frac{\Pr\{p_d \in (a, b), q_d \in (c, d) \mid D\}}{(b-a) \times (d-c)}; 0 < a < b < 1; 0 < c < d < 1. \tag{3.3}
\]

Here, the numerator, \(\Pr\{p_d \in (a, b), q_d \in (c, d) \mid D\}\), is the posterior probability of \(p_d\) and \(q_d\) falling in the interval \((a, b)\) and \((c, d)\), respectively.

Assume the prior for each \(p_d\) follows an independent \(\text{beta}(a_1, b_1)\), and the prior for each \(q_d\) follows an independent \(\text{beta}(a_2, b_2)\), where \(\text{beta}(\alpha, \beta)\) denotes a beta distribution with mean \(\frac{\alpha}{\alpha + \beta}\). The posterior distributions for \(p_d\) and \(q_d\) are \(\text{beta}(a_1 + x_d, b_1 + n_d - x_d)\) and \(\text{beta}(a_2 + y_d, b_2 + n_d - y_d)\), respectively.

Based on the posterior distributions, there exists a winning interval combination \((a^*, b^*) \times (c^*, d^*)\) that achieves the maximum JUPM among all the combinations in Table 3.4, and the corresponding decision for that combination is selected for treating the next cohort of patients.

The basic dose-finding concept of TEPI is as follows. Assume that the current patient cohort is treated at dose \(d\). After the current cohort completes DLT and response evaluation, compute the JUPMs for all the interval combinations in Table 3.4. The TEPI design recommends E,” S,” or D”, Your trial designs anywhere, anytime. 137
corresponding to the combination with the largest JUPM value according to Table 3.4.

In practice, the TEPI design needs to be calibrated according to physicians’ needs. This is transparent and requires some effort. The tuning is for the intervals in Table 3.4 so that the dosing decisions are satisfactory to the clinicians.

To enable ethical constraints, below are two additional rules as part of the dose-finding algorithm to exclude any dose with excessive toxicity and any dose with unacceptable efficacy.

**Safety and futility rules**

- **Safety rule:** if \( \Pr(p_d > p_T \mid x_d, n_d) > p_{cut} \) for a \( p_{cut} \) close to 1 (say, 0.95), exclude doses \( d, d + 1, \cdots, D \), from future use in the trial; treat the next cohort of patients at dose \( d - 1 \).

- **Futility rule:** if \( \Pr(q_{d} < q_E \mid y_d, n_d) > q_{cut} \) for a \( q_{cut} \) close to 1 (say, 0.7), where \( q_E \) is the minimum acceptable probability of efficacy, then exclude dose \( d \) from future use in the trial. Here, \( q_E \) is the reference efficacy rate, e.g., the efficacy rate of standard care.

Note that, here we assume the prior for each \( p_d \) follows an independent \( \text{beta}(a_1, b_1) \), and the prior for each \( q_d \) follows an independent \( \text{beta}(a_2, b_2) \), where \( \text{beta}(\alpha, \beta) \) denotes a beta distribution with mean \( \alpha/(\alpha + \beta) \). The posterior distributions for \( p_d \) and \( q_d \) in the above rules are \( \text{beta}(a_1 + x_d, b_1 + n_d - x_d) \) and \( \text{beta}(a_2 + y_d, b_2 + n_d - y_d) \), respectively.

**Stopping rules**

The trial is stopped if

1. the prespecified maximum total sample size \( n \) is reached; or
2. the lowest dose shows excessive toxicity according to the safety rule; In this case, the trial is early stopped and the MTD cannot be determined; or
3. **optional:**
   - the TEPI decision is “S”, to stay at the current dose, and the current dose has enrolled \( K \) patients;
   - the TEPI decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled \( K \) patients;
   - the TEPI decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled \( K \) patients.
3.3.2.3 Dose Selection

At the end of the trial, TEPI selects the most desirable dose as the OBD based on a utility score that balances the toxicity and efficacy trade-off. The utility score function is defined as $U(p, q) = f_1(p)f_2(q)$ (suppressing dose $d$ in the notation), where $p$ denotes the toxicity rate, and $q$ denotes the efficacy rate.

Both $f_1(\cdot)$ and $f_2(\cdot)$ are truncated linear functions, given by

$$f_1(p) = \begin{cases} 
1, & p \in (0, p^*_1], \\
1 - \frac{p - p^*_1}{p^*_2 - p^*_1}, & p \in (p^*_1, p^*_2), \\
0, & p \in [p^*_2, 1)
\end{cases}$$

(3.4)

$$f_2(q) = \begin{cases} 
0, & q \in (0, q^*_1], \\
\frac{q - q^*_1}{q^*_2 - q^*_1}, & q \in (q^*_1, q^*_2), \\
1, & q \in [q^*_2, 1)
\end{cases}$$

(3.5)

where $p^*$'s and $q^*$'s are prespecified cutoff values. The OBD is selected according to the following process.

1. We generate a total of $T$ random samples, \{p^{(t)}_d, t = 1, \cdots, T\} and \{q^{(t)}_d, t = 1, \cdots, T\}, from the posterior distributions $\text{beta}(a_0 + x_d, b_0 + n_d - x_d)$ and $\text{beta}(a_0 + y_d, b_0 + n_d - y_d)$ for each dose $d$, respectively. Here, U-Design sets $a_0 = b_0 = 0.005$ and $T = 1000$.

2. For toxicity probabilities of all doses in each sample $t$, $p^{(t)} = (p^{(t)}_1, \cdots, p^{(t)}_D)$, we perform isotonic transformation using the pool adjacent violators algorithm (PAVA; Mair et al. 2009) on $p^{(t)}$ to obtain $\tilde{p}^{(t)} = (\tilde{p}^{(t)}_1, \cdots, \tilde{p}^{(t)}_d)$, where $\tilde{p}^{(t)}_i \leq \tilde{p}^{(t)}_j$ if $i < j$.

3. Let $\mathcal{A} = \{d \mid n_d > 0, d = 1, \cdots, D\}$ denote the candidate dose set from which doses have been excluded according to safety and futility rules, the final selected dose $d^*$ is the one that maximizes utility scores $U(p_d, q_d)$, that is, $d^* = \arg\max_{d \in \mathcal{A}} E[U(p_d, q_d) \mid Data]$, where

$$E[U(p_d, q_d) \mid Data] = \frac{1}{T} \sum_{t=1}^{T} U(\tilde{p}^{(t)}_d, q^{(t)}_d).$$
3.3.3 The Probability Intervals of Toxicity and Efficacy (PRINTE) Design

PRINTE (Lin and Ji, 2020a) building upon previous work in TEPI (Li et al., 2017), is a dose-finding design which utilizes both toxicity and efficacy in making dosing decisions. Similar to TEPI, PRINTE partitions the unit intervals \((0, 1)\) for both the toxicity probability \(p_i\) and efficacy probability \(q_i\) into subintervals, and makes dosing-decisions based on the posterior probability of the interval combinations. Compared to TEPI, it does not require a physician-elicited decision table, the choice of which could be arbitrary and difficult, and might be subjective to Ockhams razor (Guo et al., 2017). Instead, PRINTE utilizes a decision principle that is simple and transparent, and is commonly applied in practice.

3.3.3.1 Probability Model

Consider \(D\) ascending doses in a single-agent ACT phase I trial. Due to ethical considerations, it is always assumed that the toxicity probability \(p_d\) increases with dose level \(d\), that is, \(p_1 \leq \cdots \leq p_D\). However, the efficacy probability \(q_d\) may increase initially and then reach a plateau from which minimal improvement or even decreasing efficacy may be seen with increasing dose. For this reason, we assume that \(q_d\) is not monotone with \(d\), and that \(p_d\) and \(q_d\) are independent. Suppose that dose \(d\) is currently used in the trial and \(n_d\) patients have already been allocated to dose \(d\), with \(x_d\) and \(y_d\) patients experiencing toxicity and efficacy outcomes. Aggregating across all the doses, the trial data are denoted as \(Data = \{(n_d, x_d, y_d), d = 1, \cdots, D\}\).

Let \(p_T\) be the target toxicity probability and \(p_E\) be the target efficacy rate. Define the equivalence interval (EI) as \([p_T - \epsilon_1, p_T + \epsilon_2]\) where \(\epsilon_1\) and \(\epsilon_2\) are two small fractions that allow toxicity probability of MTD to be in a range of values, rather than a single point \(p_T\).

Consider the unit square of \(Q = (0, 1) \times (0, 1)\) (here, operation \(\times\) represents the Cartesian product) representing the joint probability square of toxicity and efficacy probabilities. For toxicity, there are three probability intervals, \((0, p_T - \epsilon_1)\), \([p_T - \epsilon_1, p_T + \epsilon_2]\), and \((p_T + \epsilon_2, 1)\), which represent the under-dosing, equivalence, and over-dosing intervals. For efficacy, consider two probability intervals, \((0, p_E)\) and \((p_E, 1)\), which corresponds to low and high probability of efficacy. Denote \(S_{\text{tox}} = \{(0, p_T - \epsilon_1), [p_T - \epsilon_1, p_T + \epsilon_2], (p_T + \epsilon_2, 1)\}\) as the set of three toxicity probability intervals and \(S_{\text{eff}} = \{(0, p_E), (p_E, 1)\}\) as the set of two efficacy probability intervals. Taking a Cartesian product of the two sets, we obtain a set of six probability rectangles (PRs) in \(Q\), which is given by

\[\text{Your trial designs anywhere, anytime.}\]
3.3. Statistical Methods Review

3.3.3. The Probability Intervals of Toxicity and Efficacy (PRINTE) Design

\[ S_{\text{joint}} = S_{\text{tox}} \times S_{\text{eff}} \]

\[ = \{ s_{lh} = (0, p_T - \epsilon_1) \times (0, p_E), s_{lh} = (0, p_T - \epsilon_1) \times (p_E, 1), \]

\[ s_{el} = [p_T - \epsilon_1, p_T + \epsilon_2] \times (0, p_E), s_{eh} = [p_T - \epsilon_1, p_T + \epsilon_2] \times (p_E, 1), \]

\[ s_{hl} = (p_T + \epsilon_2, 1) \times (0, p_E), s_{hh} = (p_T + \epsilon_2, 1) \times (p_E, 1) \}, \]

where the two letters \( l \) and \( h \) denotes low or high, respectively. See Figure 3.19a for a display of the probability rectangles in \( S_{\text{joint}} \).

Figure 3.19: An example demonstrating the 2-dimensional probability rectangles and sub-rectangles of toxicity and efficacy. (a): The horizontal axis is the probability intervals of toxicity \((0, p_T - \epsilon_1), [p_T - \epsilon_1, p_T + \epsilon_2], \) and \((p_T + \epsilon_2, 1)\). The vertical axis is the probability intervals of efficacy \((0, p_E)\) and \((p_E, 1)\). The Cartesian product of both probability intervals is shown as the 6 probability rectangles (PRs) separated by dashed lines. (b): The horizontal axis is the probability sub-intervals of toxicity, where \((0, p_T - \epsilon_1), [p_T - \epsilon_1, p_T + \epsilon_2] \) and \((p_T + \epsilon_2, 1)\) are further divided into smaller intervals with the same length of \( \epsilon_1 + \epsilon_2 \). The vertical axis is the probability sub-intervals of efficacy, where \((0, p_E)\) and \((p_E, 1)\) are further divided into multiple smaller intervals with the same length of their maximum common divisor. The Cartesian product of all probability sub-intervals is shown as the probability sub-rectangles (sub-PRs) separated by dashed lines.

Divide the six PRs into sub-PRs with similar area, see Figure 3.19b for an illustration, which is realized by three steps.

1. For the toxicity interval set \( S_{\text{tox}} \), divide \( S_{\text{tox}} \) into sub-intervals given by the length of the
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

equivalence interval \((\epsilon_1 + \epsilon_2)\). The division is done by keeping the equivalence interval \(m_e^t = [p_T - \epsilon_1, p_T + \epsilon_2]\) unchanged, and sub-divide the under-dosing interval \((0, p_T - \epsilon_1)\) and over-dosing interval \((p_T + \epsilon_2, 1)\) into sub-intervals with the length \(l_e = \epsilon_1 + \epsilon_2\), except for the sub-intervals on the boundary. Denote the set of all the resulting sub-intervals as \(M_{tox} = \{m_{l,s}^t, m_{e}^t, m_{h,s}^t\}\), in which \(m_{l,s}^t\) and \(m_{h,s}^t\) are the sub-intervals generated by dividing the under-dosing and over-dosing intervals, respectively.

2. For the efficacy interval set \(S_{eff}\), divide two intervals in \(S_{eff}\) into sub-intervals with the length \(l_e\),
\[
    l_e = \max \left\{ 0.10, \frac{\gcd(100 \cdot p_E, 100 \cdot (1 - p_E))}{100} \right\},
\]
where \(\gcd(a, b)\) is the greatest common divisor of \(a\) and \(b\). Denote the resulting set of sub-intervals by \(M_{eff} = \{m_{l,s}^e, m_{h,s}^e\}\), where
\[
    m_{l,s}^e = \{(0, p_E - t_1 l_e], \cdots, (p_E - 2l_e, p_E - l_e], (p_E - l_e, p_E]\),
\]
\[
    m_{h,s}^e = \{(p_E, p_E + l_e), (p_E + l_e, p_E + 2l_e), \cdots, (p_E + 2t_2 l_e, 1]\}.
\]
Here, \(t_1\) and \(t_2\) are the maximum positive integers such that \(p_E - t_1 l_e > 0\) and \(p_E + t_2 l_e < 1\), respectively.

3. Take Cartesian product of the set of \(M_{tox}\) and \(M_{eff}\) to generate a set of two-dimensional sub-PRs of equal area, except for those on the boundary of the toxicity axis next to 0 or 1. These sets are denoted by \(M_{joint}\) as illustrated below, where \(k_{uv}, u \in \{l, e, h\}, v \in \{l, h\}\) denotes the number of sub-PRs in \(m_{uv}\).
\[
    M_{joint} = M_{tox} \times M_{eff}
    = \{m_{l,s}^t, m_{e}^t, m_{h,s}^t\} \times \{m_{l,s}^e, m_{h,s}^e\}
    = \{m_{l}^l = \{m_{l1}^1, \cdots, m_{l1}^{k_{l1}}\}, m_{lh} = \{m_{lh1}, \cdots, m_{lh}^{k_{lh1}}\},\)
    m_{el} = \{m_{el1}, \cdots, m_{el}^{k_{el1}}\}, m_{eh} = \{m_{eh1}, \cdots, m_{eh}^{k_{eh1}}\},\)
    m_{pl} = \{m_{pl1}, \cdots, m_{pl}^{k_{pl1}}\}, m_{ph} = \{m_{ph1}, \cdots, m_{ph}^{k_{ph1}}\}\}
\]

PRINTe treats each sub-PR as a model and considers a model indicator \(a\) that takes one of the sub-PRs. Denote \(m_{uv}\) as a sub-PR in the set \(M_{joint}\), and define \(\{a = m_{uv}\} = \{(p_d, q_d) \in m_{uv}\}\). Embedding the model indicator \(a\) into a Bayesian hierarchic model, we compute the posterior probability of each sub-PR given the observed toxicity and efficacy outcomes \(\{x_d, y_d\}\), given by
\[
    P(a = m_{uv} \mid x_d, y_d, n_d) = Pr((p_d, q_d) \in m_{uv} \mid x_d, y_d, n_d).\]
From model selection perspective,
finding the optimal decision is equivalent to selecting the optimal model (sub-PR) that maximizes the marginal posterior model probability.

We further define dose-finding decisions as \( a^* \in \{ E, S, D \} \) and maps \( a \in \{m_{ll}, m_{lh}, m_{el}, m_{eh}, m_{hl}, m_{hh} \} \) to \( a^* \in \{ E, S, D \} \) according to the following rule \( \mathcal{R} \).

\[
\begin{align*}
    a^* = \mathcal{R}(a) &= \begin{cases} 
        E, & \text{if } a = m_{ll} \\
        E, & \text{if } a = m_{el} \text{ and } n_{d+1} = 0 \\
        S, & \text{if } a = m_{el} \text{ and } n_{d+1} > 0 \\
        S, & \text{if } a \in \{m_{lh}, m_{eh}\} \\
        D, & \text{if } a \in \{m_{hl}, m_{hh}\}
    \end{cases}
\end{align*}
\]

The rule \( \mathcal{R} \) states that the dosing decisions \( \{ E, S, D \} \) correspond to the models that describe the toxicity and efficacy probabilities of the dose. According to \( \mathcal{R}(a) \), escalation (\( E \)) is recommended if toxicity and efficacy are both deemed low; Stay (\( S \)) is selected if \( n_{d+1} > 0 \), toxicity is near the MTD range and efficacy is low, while escalation (\( E \)) is recommended if \( n_{d+1} = 0 \), i.e., dose \((d + 1)\) is untried; Stay (\( S \)) is selected if either 1) toxicity is low but efficacy is high \( m_{lh} \), or 2) toxicity is near the MTD range and efficacy is high; Lastly, de-escalation \( D \) is selected if toxicity is high regardless of efficacy. The goal is to seek an optimal \( a \) that leads to an optimal decision \( a^* \).

### 3.3.3.2 Dose-finding Algorithm

The implementation of PRINTE is simple and transparent. The only required input values are \( p_T, p_E \), and the equivalence interval \([p_T - \epsilon_1, p_T + \epsilon_2]\). Once they are provided, optimal decisions \( a^{opt\star} \) can be calculated for all possible toxicity and efficacy outcomes at a given dose. Suppose that the current dose is \( d, d \in \{1, ..., D\} \). Record \( \{x_d, y_d, n_d\} \) and calculate the marginal model posterior probabilities \( \Pr(a \mid x_d, y_d, n_d) \), and then the optimal decision \( a^{opt\star} \) can be determined. The next cohort of patients is allocated to \( \{\max(1, d - 1), d, \min(d + 1, D)\} \) according to \( a^{opt\star} \).

**Safety and futility rules**

- **Safety rule**: if \( \Pr(p_d > p_T \mid x_d, n_d) > p_{cut} \) for a \( p_{cut} \) close to 1 (say, 0.95), exclude doses \( d, d + 1, \cdots, D \), from future use in the trial; treat the next cohort of patients at dose \((d - 1)\).

- **Futility rule**: if \( \Pr(q_d < q_E \mid y_d, n_d) > q_{cut} \) for a \( q_{cut} \) close to 1 (say, 0.7), where \( q_E \) is the minimum acceptable probability of efficacy, then exclude dose \( d \) from future use in the trial. Here, \( q_E \) is the reference efficacy rate, e.g., the efficacy rate of standard care.
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

Note that, here we assume the prior for each $p_d$ follows an independent $\text{beta}(a_1, b_1)$, and the prior for each $q_d$ follows an independent $\text{beta}(a_2, b_2)$, where $\text{beta}(\alpha, \beta)$ denotes a beta distribution with mean $\alpha/(\alpha + \beta)$. The posterior distributions for $p_d$ and $q_d$ in the above rules are $\text{beta}(a_1 + x_d, b_1 + n_d - x_d)$ and $\text{beta}(a_2 + y_d, b_2 + n_d - y_d)$, respectively.

**Stopping rules**

The trial is stopped if

1. the prespecified maximum total sample size $n$ is reached; or
2. the lowest dose shows excessive toxicity according to the safety rule; In this case, the trial is early stopped and the MTD cannot be determined; or
3. optional:
   - the PRINTE decision is “S”, to stay at the current dose, and the current dose has enrolled $K$ patients;
   - the PRINTE decision is “E”, to escalate to the next higher dose, and that next higher dose has enrolled $K$ patients;
   - the PRINTE decision is “D”, to de-escalate to the previous lower dose, and that previous lower dose has enrolled $K$ patients.

### 3.3.3.3 Dose Selection

At the end of the trial, PRINTE chooses the OBD using a joint utility score $U(p, q) = f_1(p)f_2(q)$ (suppressing dose $d$ in the notation), which takes the product of toxicity utility $f_1(p)$ in (3.6) and efficacy utility $f_2(q)$ in (3.7).

\[
f_1(p) = \begin{cases} 
1, & p \in (0, p_1^*), \\
1 - \frac{p - p_1^*}{p_2^* - p_1^*}, & p \in (p_1^*, p_2^*), \\
0, & p \in (p_2^*, 1). 
\end{cases} \tag{3.6}
\]

\[
f_2(q) = \begin{cases} 
0, & q \in (0, q_1^*), \\
\frac{q - q_1^*}{q_2^* - q_1^*}, & q \in (q_1^*, q_2^*), \\
1, & q \in (q_2^*, 1). 
\end{cases} \tag{3.7}
\]

For toxicity, define two thresholds $p_1^*$ and $p_2^*$ such that the toxicity utility score is 1 when $p < p_1^*$, 0 when $p > p_2^*$, and linearly decreases when $p$ is between $(p_1^*, p_2^*)$. For efficacy, define...
two thresholds $q^*_1$ and $q^*_2$ such that the efficacy utility score is 0 when $q < q^*_1$, is 1 when $q > q^*_2$, and linearly increases when $q$ is between $(q^*_1, q^*_2)$. The OBD is selected according to the following process.

1. We generate a total of $T$ random samples, \( \{p^{(t)}_d, t = 1, \ldots, T\} \) and \( \{q^{(t)}_d, t = 1, \ldots, T\} \), from the posterior distributions \( \text{beta}(a_0 + x_d, b_0 + n_d - x_d) \) and \( \text{beta}(a_0 + y_d, b_0 + n_d - y_d) \) for each dose $d$, respectively. Here, U-Design sets $a_0 = b_0 = 0.005$ and $T = 1000$.

2. For toxicity probabilities of all doses in each sample $t$, \( p^{(t)} = (p^{(t)}_1, \ldots, p^{(t)}_D) \), we perform isotonic transformation using the pool adjacent violators algorithm (PAVA; Mair et al. 2009) on $p^{(t)}$ to obtain \( \tilde{p}^{(t)} = (\tilde{p}^{(t)}_1, \ldots, \tilde{p}^{(t)}_d) \), where $\tilde{p}^{(t)}_i \leq \tilde{p}^{(t)}_j$ if $i < j$.

3. We propose a probabilistic inference for selecting the OBD and avoid selecting doses with low utility. Define an admissible probability region (APR) \( A(p, q) = \{(p, q) \mid p \in (0, p_f), q \in [q_E, 1]\} \). Then the OBD is selected only from the candidate dose set $\mathcal{A}$,

\[
\mathcal{A} = \{d \mid p_{\text{in},d} \geq p_{\text{grad}}, n_d > 0, d = 1, \ldots, D\},
\]

where $p_{\text{in},d} = \Pr\{(p_d, q_d) \in \text{APR} \mid \text{Data}\}$ is the posterior probability that dose $d$ belongs to APR and $p_{\text{grad}}$ is a small value (say, 0.1). We use a simple a simple numerical approximation approach to compute $p_{\text{in},d}$ given by

\[
\hat{p}_{\text{in},d} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{ (\tilde{p}^{(t)}_d, q^{(t)}_d) \in \text{APR} \}.
\]

4. The final selected dose $d^*$ is the one that maximizes the utility score $U(p_d, q_d)$. That is, $d^* = \arg\max_{d \in \mathcal{A}} \hat{E}[U(p_d, q_d) \mid \text{Data}]$, where

\[
\hat{E}[U(p_d, q_d) \mid \text{Data}] = \frac{1}{T} \sum_{t=1}^{T} U(\tilde{p}^{(t)}_d, q^{(t)}_d).
\]
3.3.4 The Efficacy-Toxicity (EffTox) Trade-Offs-Based Design

EffTox, proposed in Thall and Cook (2004), is an outcome-adaptive, model-based Bayesian procedure that chooses doses of an experimental agent for successive patient cohorts in a clinical trial based on both efficacy (E) and toxicity (T) outcomes. EffTox models the dose-efficacy and dose-toxicity relationship respectively using two different dose-response curves. Based on accumulating efficacy and toxicity data over the trial, EffTox continuously updates the parameters of the dose-response models. The desirability of each dose \( x \) is evaluated by using a family of contours characterizing the trade-off between E and T, and patients are assigned to the most desirable dose in cohorts.

3.3.4.1 Dose-Outcome Models

Assume \( D \) dose \( s_1, \ldots, s_D \) to be considered in the trial, and code dose as

\[
x_d = \log(s_d) - D^{-1} \sum_{k=1}^{D} \log(s_k)
\]  

(3.8)

for use in the regression models. If \( 0 = s_1 < s_2 \), first add \( s_2 \) to each \( s_d \) before taking logs. Let \( \pi(x, \theta) = \{ \pi_E(x, \theta), \pi_T(x, \theta) \} \) be the probabilities of efficacy and toxicity, where \( x \) denotes dose and \( \theta \) is the model parameter vector.

Given the current interim trial data \( D \), define \( x \) to be an acceptable dose if

\[
\Pr\{\pi_E(x, \theta) > q_E \mid D\} > 1 - q_{\text{cut}}
\]

(3.9)

and

\[
\Pr\{\pi_T(x, \theta) < p_T \mid D\} > 1 - p_{\text{cut}},
\]

(3.10)

where \( q_E \) and \( p_T \) are fixed lower and upper limits specified by the physician, and \( q_{\text{cut}} \) and \( p_{\text{cut}} \) are fixed probability cutoffs.

For toxicity, assume logit\((\pi_T(x, \theta)) = \mu_T + x \beta_T \), in which we set \( \beta_T > 0 \) to meet the monotonic dose-toxicity assumption. For efficacy, to allow a wide variety of possible doseresponse relationships, assume logit\((\pi_E(x, \theta)) = \mu_E + x \beta_{E,1} + x^2 \beta_{E,2} \). For simplicity, temporarily suppress \((x, \theta)\). The joint outcome model is given by

\[
\pi_{a,b} = (\pi_E)^a (1 - \pi_E)^{1-a} (\pi_T)^b (1 - \pi_T)^{1-b} + (-1)^{a+b} \pi_E (1 - \pi_E) \pi_T (1 - \pi_T) \left( e^{\psi} - 1 \right) \left( e^{\psi} + 1 \right)
\]

(3.11)

for \( a, b \in \{0, 1\} \) and real-valued \( \psi \). Thus, \( \theta = (\mu_T, \beta_T, \mu_E, \beta_{E,1}, \beta_{E,2}, \psi) \). Since \( \beta_T \) should be greater than 0, we assume that \( \beta_T \) is lognormally distributed, with mean \( \bar{\mu}_{\beta_T} \) and standard deviation
3.3.4. The EfficacyToxicity (EffTox) Trade-Offs-Based Design

To determine the desirability of each dose, the EffTox design constructs a efficacy-toxicity desirability contour, \( C \), in the two-dimensional domain \( \Pi = [0, 1]^2 \) by fitting a curve to target values of \( \pi \) elicited from the physician. The contour \( C \) is then used to construct a family of desirability contours such that all \( \pi \) on the same contour are equally desirable. Because the family of contours partitions \( \Pi \), this construction provides a basis for comparing doses in terms of their posterior means, \( E\{\pi(x, \theta) \mid D\} \).

To construct \( C \), we first elicit three target values, \( \{\pi_1^*, \pi_2^*, \pi_3^*\} \), which the physician considers equally desirable. First, elicit a desirable trade-off target, \( \pi_1^* = (\pi_{1,E}^*, \pi_{1,T}^*) = (\pi_{1,E}^*, 0) \), in the case where toxicity has probability 0. That is, elicit the smallest efficacy probability, \( \pi_{1,E}^* \), that the physician would consider desirable if toxicity were impossible. Next, elicit \( \pi_2^* \) having the same desirability as \( \pi_1^* \) by asking the physician what the maximum value of \( \pi_T \) may be if \( \pi_E = 1 \). Given these two equally desirable extremes, elicit a third pair, \( \pi_3^* \), that is equally desirable but is intermediate between \( \pi_1^* \) and \( \pi_2^* \).

The desirability function of \( (\pi_E, \pi_T) = \pi \in [0, 1]^2 \) is defined to be

\[
\delta(\pi_E, \pi_T) = 1 - \| (\pi_E, \pi_T) - (1, 0) \|_p \\
= 1 - \left\{ \left( \frac{\pi_E - 1}{\pi_{1,E}^* - 1} \right)^p + \left( \frac{\pi_T - 0}{\pi_{2,T}^* - 0} \right)^p \right\}^{1/p}
\]

(3.12)

where \( p > 0 \). Solve \( \delta(\pi_E^*, \pi_T^*) = 0 \) for \( p \) using the bisection method, wherein intervals known to bracket the solution are successively refined (Peter et al., 2014). This gives \( \delta(\pi) = 0 \) on \( C \) with \( \delta(\pi) \) increasing as \( \pi \) moves along any straight line from a point in \([0, 1]^2\) to the ideal pair \( (\pi_E, \pi_T) = (1, 0) \). After solving for \( p \), the desirability measure can be computed for any point \( (\pi_E, \pi_T) \) using formula (3.12).

The following definition exploits this structure to induce an ordering on the set of doses.

**DEFINITION:** Given \( D \) and \( x \), the desirability, \( \delta(x, D) \), of \( x \) is the desirability of the posterior mean \( E\{\pi(x, \theta) \mid D\} \).

*Your trial designs anywhere, anytime.*
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

Figure 3.20: Example of efficacy-toxicity desirability contours. The contour $C$ is the line with desirability equals to 0 ($U = 0.0$).

To apply this during the trial, after the most recent cohorts data have been incorporated into $D$, for each $x, (\pi_E, \pi_T) = E\{\pi(x, \theta) | D\}$ is first computed, and then the desirability of $x$ is computed by formula (3.12). Among the doses with acceptable efficacy and toxicity, the dose that maximizes $\delta(x, D)$ is selected.

3.3.4.3 The Trade-Off-Based Algorithm

Initially, the physician must provide a set of doses, a starting dose for the first cohort, $N$, $c$, and the limits $q_E$ and $p_T$ used in the acceptability criteria (3.9) and (3.10). The trade-off targets $\{\pi_1^*, \pi_2^* \pi_3^*\}$ then must be elicited in order to construct $C$ and the family of trade-off contours. The probability cut-offs $q_{cut}$ and $p_{cut}$ in (3.9) and (3.10) are determined, using preliminary computer simulation results, to obtain a design with desirable operating characteristics. Given this structure, the dose-finding algorithm proceeds as follows:

1. Treat the first cohort at the starting dose specified by the physician.
2. For each cohort after the first, $x \in A(D)$ if $x$ satisfies both (3.9) and (3.10), or if $x$ is the lowest untried dose above the starting dose and it satisfies (3.10).
3. If $A(D) \neq \phi$, then the next cohort is treated at the most desirable $x \in A(D)$, subject to the constraint that no untried dose may be skipped when escalating.
4. If $A(D) = \phi$, then the trial is terminated and no dose is selected.
5. If the trial is not stopped early and $A(D_N) \neq \phi$ at the end of the trial, then the dose $x \in A(D_N)$...
3.3. Statistical Methods Review

3.3.4. The Efficacy-Toxicity (EffTox) Trade-Offs-Based Design

\[ A(D_N) \text{ maximizing } \delta(x, D_N) \text{ is selected.} \]
3.3.5 The Utility-Based Bayesian Optimal Interval (U-BOIN) Design

U-BOIN (Zhou et al., 2019) is a model-based design that jointly models toxicity and efficacy using a multinomial-Dirichlet model and employ a utility function to measure dose risk-benefit trade-off. The design consists of two seamless stages. In stage I, the Bayesian optimal interval (BOIN) design (Liu and Yuan, 2015b) is used to quickly explore the dose space and collect preliminary toxicity and efficacy data. In stage II, the posterior estimate of the utility for each dose is continuously updated using accumulating efficacy and toxicity data, and the posterior estimate is used to direct patient allocation and OBD selection.

3.3.5.1 Efficacy-Toxicity Model

Consider a phase I/II trial with $J$ doses under investigation. Let $Y_E$ denote the binary efficacy endpoint, where $Y_E = 1$ denotes response, and 0 otherwise; let $Y_T$ denote the binary toxicity endpoint, where $Y_T = 1$ denotes DLT, and 0 otherwise. The bivariate discrete outcome $(Y_E, Y_T)$ can be equivalently represented by a single variable $Y$ with $2 \times 2 = 4$ levels, with $Y = 1$, if $(Y_E, Y_T) = (0, 1)$; $Y = 2$, if $(Y_E, Y_T) = (0, 0)$; $Y = 3$, if $(Y_E, Y_T) = (1, 1)$; and $Y = 4$, if $(Y_E, Y_T) = (1, 0)$. Here $Y = 1$ is the least favorable clinical outcome (DLT, no efficacy), and $Y = 4$ denotes the most favorable clinical outcome (No DLT, efficacy).

Define $\pi_{jk} = Pr(Y = k \mid d = j)$, where $k = 1, \ldots, 4$ and $j = 1, \ldots, J$, with $\sum_{k=1}^{4} \pi_{jk} = 1$, where $d$ denotes the dose level. Assume that $Y$ follows a Dirichlet-multinomial model as follows:

$$Y = k \mid d = j \sim \text{Multinomial}(\pi_{j1}, \ldots, \pi_{j4}) \quad (3.13)$$

$$(\pi_{j1}, \ldots, \pi_{j4}) \sim \text{Dirichlet}(a_1, \ldots, a_4) \quad (3.14)$$

where $a_1, \ldots, a_4 > 0$ are hyperparameters. U-Design sets $a_k = \frac{1}{4}$, $k = 1, \ldots, 4$, as the default values, such that the prior is vague and equivalent to an effective sample size of 1.

Assume that $n_j$ patients have been treated at dose $d = j$, among whom $n_{jk}$ patients had outcome $Y = k$. Denote $D_j = (n_{j1}, \ldots, n_{j4})$, and the posterior distribution of $\pi_j = (\pi_{j1}, \ldots, \pi_{j4})$ is

$$\pi_j \mid D_j \sim \text{Dirichlet}(a_1 + n_{j1}, \ldots, a_4 + n_{j4}). \quad (3.15)$$

3.3.5.2 Utility

Let $\psi_k$ denote the utility value ascribed to outcome $Y = k$, $k = 1, \ldots, 4$, which can be elicited from physicians to reflect the risk-benefit trade-off underlying their medical decisions using the following
3.3. Statistical Methods Review

3.3.5. The Utility-Based Bayesian Optimal Interval (U-BOIN) Design

- Fix the value of the utility for the least desirable outcome $Y = 1$ as $\psi_1 = 0$, and for the most desirable outcome $Y = 4$ as $\psi_4 = 1$.
- Ask the clinician to use these two utilities as a reference to score the utility values $\psi_2, \psi_3$ for the other 2 possible outcomes $Y = 2, 3$ to quantify the risk-benefit trade-off under each outcome.

Table 3.5 shows two examples of the utility function.

<table>
<thead>
<tr>
<th>(a) Example 1</th>
<th>(b) Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_T = 1$</td>
<td>$Y_T = 1$</td>
</tr>
<tr>
<td>$Y_E = 0$</td>
<td>$Y_E = 0$</td>
</tr>
<tr>
<td>$\psi_1 = 0$</td>
<td>$\psi_1 = 0$</td>
</tr>
<tr>
<td>$\psi_2 = 0.3$</td>
<td>$\psi_2 = 0.3$</td>
</tr>
<tr>
<td>$Y_T = 0$</td>
<td>$Y_T = 0$</td>
</tr>
<tr>
<td>$Y_E = 1$</td>
<td>$Y_E = 1$</td>
</tr>
<tr>
<td>$\psi_3 = 0.5$</td>
<td>$\psi_3 = 0.65$</td>
</tr>
<tr>
<td>$\psi_4 = 1$</td>
<td>$\psi_4 = 1$</td>
</tr>
</tbody>
</table>

Example 1 has utility values $\{\psi_1 = 0, \psi_2 = 0.3, \psi_3 = 0.5, \psi_4 = 1\}$ for the outcomes $\{(Y_E = 0, Y_T = 1), (Y_E = 0, Y_T = 0), (Y_E = 1, Y_T = 1), (Y_E = 1, Y_T = 0)\}$, respectively. Compared to example 1, example 2 rewards the response (i.e., $Y_E = 1$) more, in the presence of DLT (i.e., $Y_T = 1$), by assigning a larger value to $\psi_3$ (0.65 versus 0.50). This is appropriate for a trial where toxicity can be well managed and efficacy response is highly desirable (e.g., leading to long survival).

Given the values of $\psi_k$, the true mean utility for dose $j$ is given by

$$U_j = \sum_{k=1}^{4} \psi_k \pi_{jk}.$$

Since the true mean utility $U_j$ depends on $\pi_{jk}$, which is unknown, it is estimated based on the observed data. Given the interim data $D = \{D_j\}$, the estimate of mean utility is given by

$$\hat{U}_j = \sum_{k=1}^{4} \psi_k E(\pi_{jk} \mid D).$$

3.3.5.3 Optimal Biological Dose

Let $p_T$ denote the maximum tolerable DLT rate, and $q_E$ the lowest acceptable response rate. Let $\pi_{T,j} = \pi_{j1} + \pi_{j3} = Pr(Y_T = 1 \mid d = j)$ and $\pi_{E,j} = \pi_{j3} + \pi_{j4} = Pr(Y_E = 1 \mid d = j)$. Define
Module 3. Single-Agent Dose-Finding Designs with Efficacy & Toxicity Endpoints and Cohort Enrollment

that dose $j$ is inadmissible, if it meets either one or both of the following two criteria:

$$\Pr(\pi_{T,j} > p_T | D) > p_{cut} \tag{3.18}$$

$$\Pr(\pi_{E,j} < q_E | D) > q_{cut} \tag{3.19}$$

where $p_{cut}$ and $q_{cut}$ are probability cutoffs. According to (3.13) and (3.14), $\pi_{T,j}$ and $\pi_{E,j}$ follow posterior beta distributions, given by

$$\pi_{T,j} | D \sim \text{Beta}(a_1 + a_3 + n_{j_1} + n_{j_3}, a_2 + a_4 + n_{j_2} + n_{j_4}),$$

$$\pi_{E,j} | D \sim \text{Beta}(a_3 + a_4 + n_{j_3} + n_{j_4}, a_1 + a_2 + n_{j_1} + n_{j_2}).$$

The admissible dose is then defined as the dose for which none of the criteria (3.18) and (3.19) is satisfied. Define the OBD as the dose that is admissible and has the highest utility value, i.e.,

$$\text{OBD} = \arg \max_{j \in \mathcal{A}} (U_j) \tag{3.20}$$

where $\mathcal{A}$ denotes the set of admissible doses.

3.3.5.4 Dose-finding Algorithm

The U-BOIN design consists of two seamless stages (Figure 3.21). The objective of stage I is to quickly explore the dose space to identify a set of admissible doses that are reasonably efficacious and safe for stage II. In stage I, dose escalation is conducted based on only the toxicity outcome. However, efficacy data are also collected and will be used for decision making in stage II. Stage I dose escalation/de-escalation is guided by the BOIN design (Liu and Yuan, 2015b). Due to very limited data and large uncertainty, for patient safety, set the target DLT rate $\phi_T = p_T - 0.05$, slightly lower than the maximum tolerable DLT rate $p_T$, to ensure that stage I dose exploration concentrates around up to, but not exceeding $p_T$. Let $\hat{\pi}_{T,j}$ denote the empirical (or maximum likelihood) estimate of $\pi_{T,j}$, given by $\hat{\pi}_{T,j} = \frac{m_j}{n_j}$ where $m_j$ is the number of patients who experienced DLT at the dose level $j$; and let $\lambda_e$ and $\lambda_d$ denote the predetermined optimal escalation boundary and de-escalation boundary. Table 3.6 provides the values of $\lambda_e$ and $\lambda_d$ for the commonly used target DLT rate $\phi_T$. See the work of Liu and Yuan (2015b) for the derivation and formula to calculate $\lambda_e$ and $\lambda_d$. The dose-finding algorithm in stage I proceeds as follows.

Ia. Patients in the first cohort are treated at dose level 1 or a prespecified starting dose.

Ib. Suppose $j$ is the current dose; use the following rules to assign a dose to the next cohort of patients:
3.3. Statistical Methods Review

3.3.5. The Utility-Based Bayesian Optimal Interval (U-BOIN) Design

- Escalate the dose to $j + 1$ if $\hat{\pi}_{T,j} \leq \lambda_e$.
- De-escalate the dose to $j - 1$ if $\hat{\pi}_{T,j} \geq \lambda_d$.
- Otherwise, stay at the current dose $j$.

Ic. Repeat step Ib until the number of patients treated on one of the doses reaches $s_1$, and then move to stage II.

In stage I, following the BOIN design, if $\Pr(\pi_{T,j} > p_T \mid m_j, n_j) > 0.95$ and $n_j \geq 3$, dose level $j$ and higher are eliminated from the trial; the trial is terminated if the lowest dose level is eliminated, where $\Pr(\pi_{T,j} \geq p_T \mid m_j, n_j) > 0.95$ is evaluated based on a beta-binomial model with the uniform prior.

Stage II proceeds as follows.

IIa. Let $j^*$ denote the highest dose level that has been tried. If $\hat{\pi}_{T,j^*} \leq \lambda_e$ and $j^*$ is not the highest dose in the trial, escalate the dose to $(j^* + 1)$ for treating the next cohort of patients; otherwise, proceed to step IIb.

IIb. Given the observed interim data $D$ collected in both stages I and II, determine the admissible dose set $A$ from dose 1, $\cdots$, $j^*$, where none of the criteria (3.18) and (3.19) is satisfied for each dose in $A$. If no dose is admissible, terminate the trial and no dose should be selected as the OBD. Otherwise, assign the next cohort of patients to a dose in $A$. In U-Design, there are two methods to assign the next cohort,

- Pick The Winner, assigning to dose $j \in A$ that has the largest posterior mean utility.
- Adaptive Randomization, adaptively randomizing the next cohort of patients to dose $j \in A$, with probability $\omega_j$ proportional to its posterior mean utility, i.e.,

$$\omega_j = \frac{U_j}{\sum_{j \in A} U_j}.$$

IIc. Repeat steps IIa and IIb until reaching the prespecified maximum sample size $N$ or the number of patients treated at one of the doses in stage II reach $s_2$ (Zhou et al. (2019) recommends that $s_2 > s_1$), and then select the OBD following the rules in §3.3.5.3.

Your trial designs anywhere, anytime.
Table 3.6: Dose escalation and de-escalation boundaries of the Bayesian optimal interval design

<table>
<thead>
<tr>
<th>Boundaries</th>
<th>Target DLT rate ($\hat{\phi}_T$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.15</td>
</tr>
<tr>
<td>$\lambda_e$ (escalation)</td>
<td>0.118</td>
</tr>
<tr>
<td>$\lambda_d$ (de-escalation)</td>
<td>0.179</td>
</tr>
</tbody>
</table>

Figure 3.21: Diagram of the utility-based Bayesian optimal interval (U-BOIN) design.
4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

4.1 Introduction

Combination therapy refers to the use of more than one drug in patient care and is an important therapeutics in many disease settings, including cancer, cardiovascular disease, and infectious disease. In 2013, FDA issued the guidance “Codevelopment of Two or More New Investigational Drugs for Use in Combination” (FDA, 2013), which stated that: “the use of combinations of drugs directed at multiple therapeutic targets can improve treatment response, minimize development of resistance or adverse events”. There is growing interest in the development of new investigational drug combinations.

One of the challenges in combination therapy development is to find the optimal dose of each drug when using in combination. Due to the unknown potential interactions between drugs (synergy, antagonism or no interaction), the optimal dose combination might differ from the combination of the optimal dose of each drug when used alone. In this module, we mainly pay attention to the phase I dose-finding trials in oncology, especially dose-finding trials for two agents, with the goal to capture the dose-toxicity relationship for drug combinations and to identify one or more maximum tolerated dose combination (MTDC) or a MTD contour. Only the toxicity outcome, such as dose limiting toxicity (DLT) is considered in this module. A scientific way of characterizing the drug combination-toxicity profile is to test all possible combinations of candidate dose levels of two drugs. However such an approach might be impractical because the number of combinations could be too large for an early-phase trial. For example, if two drugs are to be investigated, each with 3 dose levels, there will be a total of $3 \times 3 = 9$ possible combinations. If more than two drugs are involved, this number grows exponentially to dozens or hundreds. In practice, trialists often escalate the dose level of one drug by holding the dose of another drug at a fixed level. For example, in a
phase I trial of a newly targeted monoclonal antibody (mAb) combined with a PD-1 inhibitor, say pembrolizumab, the dose of PD-1 is often fixed at the approved level (say, 3 mg/kg) and the the dose levels of mAb are varied. If so, some single-agent dose-finding designs, such as mTPI-2 (Guo et al., 2017) and i3+3 (Liu et al., 2020), could be adopted. However, such an approach may miss the global optimal dose combination since one drug is always at a fixed dose. For example, the optimal dose level of PD-1 when administrated in combination with the mAb might be 1 mg/kg, rather than 3 mg/kg. To this end, “single-agent” dose-finding designs might not be the most scientific way to identify the dual-agents optimal dose.

How to efficiently explore the drug combination-toxicity profile is a statistical problem that requires effective modeling and decision making. In recent years, a large number of designs have been proposed to find one or more maximum tolerated dose combination (MTDC) of two agents, for example, Lyu et al. (2019); Tighiouart et al. (2017); Wages et al. (2017); Lin and Yin (2016); Wages (2016); Mander and Sweeting (2015); Neuenschwander et al. (2015); Cai et al. (2014); Riviere et al. (2014); Tighiouart et al. (2014); Wages and Conaway (2014); Shi and Yin (2013); Braun and Wang (2010); Yin and Yuan (2009); Conaway et al. (2004) etc. The MTDC is defined as the highest dose combination at which the probability that a patient experiences the DLT is closest to or less than a pre-specified target rate $p_T$, which is usually determined by physicians or clinical teams, say $p_T = 30\%$. Some of these designs have been applied to real-world trials. For example, a combination dose-finding trial (NCT02366819) uses the CI3+3 design based on the research of our team.

Here, we describe a module in U-Design, **Dual-Agents Cohort-Based Designs**, which includes the Bayesian logistic regression model (BLRM) for two agents (BLRM-2d) (Neuenschwander et al., 2015), the product of independent beta probabilities dose escalation (PIPE) design (Mander and Sweeting, 2015), and a novel design called Combo i3+3 (CI3+3).

Hereinafter, we use “drug” and “agent”, “dose” and “dose combination”, interchangeably.
4.2 User Interface and Tutorial

4.2.1 Overview

Entering the Dual-Agents Cohort-Based Designs page, users will see two main tabs: Simulation Setup and Simulation Results. These two tabs allow users to conduct simulations and visualize/download simulation results. The Simulation Setup tab requires three steps to set up simulations using one or more designs (Figure 4.1): Step 1: Set trial parameters; Step 2: Select designs; and Step 3: Generate scenarios. Upon completing steps 1-3, users click the “Launch Simulation” button at the bottom of the page. User may also click the “Reset” button to clear all settings. After the simulation is launched, the results of simulations will be displayed in the Simulation Results tab. The simulation process can be monitored in real time at the top of the Simulation Results tab. Detailed steps of using this module are elaborated in §4.2.2-§4.2.3.

![Figure 4.1: Simulation Setup in the Dual-Agents Cohort-Based Designs module.](image-url)
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

4.2.2 Simulation Setup

In the module of Dual-Agents Cohort-Based Designs, U-Design provides three designs, BLRM-2d, PIPE, and CI3+3, for simulation. Users can choose up to four design configurations for simultaneous comparison in the Simulation Setup tab each time. A design configuration means a design such as CI3+3, along with the designs settings, such as sample size. Request to allow more than four design configurations by emailing admin@laiyaconsulting.com.

4.2.2.1 Step 1: Set trial parameters

Specify the target toxicity probability ($p_T$), number of simulations ($n_{sim}$) and random seed of simulation ($R_{seed}$) for the simulation trials. See Figure 4.2. Hover mouse over each trial parameter, and a description will be displayed explaining the meaning of the parameter. The detailed description of the above three input parameters is in Table 4.1.

![Figure 4.2: Set trial parameters in the Dual-Agents Cohort-Based Designs module.](image)

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T$</td>
<td>Target toxicity probability</td>
<td>The target toxicity probability of the maximum tolerated dose (MTD). The main objective of phase I clinical trials is to find the highest dose with a toxicity probability closest to or lower than $p_T$. Default value is 0.3.</td>
</tr>
<tr>
<td>$n_{sim}$</td>
<td>The number of simulated trials</td>
<td>The maximum number of simulated trials allowed is 10,000. Default value is 1,000.</td>
</tr>
<tr>
<td>$R_{seed}$</td>
<td>The random seed of simulation</td>
<td>A random seed is a number used to initialize a pseudorandom number generator in the simulation. Default value is 32432.</td>
</tr>
</tbody>
</table>
4.2.2.2 Step 2: Select designs

To select a design, click the button with the design’s name on it. Up to four design configurations may be selected for comparison.

Click the “Delete” button to remove the selected designs.

Design’s parameters can be modified in the input box of corresponding row. Hover mouse over each design parameter, and a description will be displayed explaining the meaning of the parameter. See detailed parameter descriptions in Table 4.2.

Figure 4.3: Add designs in the Dual-Agents Cohort-Based Designs module.
Table 4.2: Input parameters for designs in the **Dual-Agents Cohort-Based Designs** module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Sample size</td>
<td>The maximum number of patients to be treated in the trial. The upper limit is set at 100 since the number of patients that are enrolled in phase I clinical trial is typically small. Default value is 30.</td>
</tr>
<tr>
<td>$n_{cohort}$</td>
<td>Cohort size</td>
<td>$s$ in each cohort. The number of patients. Default value is 3.</td>
</tr>
<tr>
<td>$\epsilon_1, \epsilon_2$</td>
<td>$\epsilon_1, \epsilon_2$</td>
<td>Two small fractions used to define the equivalence/target interval of the MTDC. Any doses with a toxicity probability falling into the interval $[p_T - \epsilon_1, p_T + \epsilon_2]$ will be considered an acceptable dose level as MTDC. Default values for both are 0.05.</td>
</tr>
<tr>
<td>$PEWOC$</td>
<td>Cutoff probability of escalation with overdose control</td>
<td>The threshold of controlling the probability of excessive or unacceptable toxicity. Default value is 0.25</td>
</tr>
<tr>
<td>$d_{start,1}$</td>
<td>Starting dose level for agent 1</td>
<td>The starting dose level for agent 1 in the simulation trials. Default value is 1.</td>
</tr>
<tr>
<td>$d_{start,2}$</td>
<td>Starting dose level for agent 2</td>
<td>The starting dose level for agent 2 in the simulation trials. Default value is 1.</td>
</tr>
</tbody>
</table>
4.2. User Interface and Tutorial
4.2.2. Simulation Setup

4.2.2.3 Step 3: Generate scenarios

Select the number of doses for two agents \( n_{\text{dose},1} \) and \( n_{\text{dose},2} \) \((2 \leq n_{\text{dose},1}, n_{\text{dose},2} \leq 5)\) from the dropdown boxes, and their dose levels, \( d_{\text{level},1} \) and \( d_{\text{level},2} \). Hover mouse over each parameter, and a description will be displayed explaining the meaning of the parameter (Figure 4.4). The default values of dose levels of agents 1 and 2 are \{1, 2, \ldots, n_{\text{dose},1}\} and \{1, 2, \ldots, n_{\text{dose},2}\}, respectively. Request to allow more than five dose levels for any agent via email admin@laiyaconsulting.com.

![Step 3: Generate scenarios](image)

**Figure 4.4:** Specify input parameters in the Generate Scenarios step of the Dual-Agents Cohort-Based Designs module.

U-Design provides four ways to generate scenarios. They are described in detail in §4.3.1. Below we provide a quick guidance.

1) automatic construction (Default Scenarios tab, see Figure 4.5),
2) logistic regression (Logistic Regression tab, see Figure 4.6),
3) specifying marginal toxicity probabilities of each agent and the interaction between two agents (Marginals & Interaction tab, see Figure 4.7),
4) manual construction (Manual Construction tab, see Figure 4.8).

1) Default Scenarios (Figure 4.5)

Upon selection of \( n_{\text{dose},1} \) and \( n_{\text{dose},2} \) and specification of \( d_{\text{level},1} \) and \( d_{\text{level},2} \), click the “Generate” button to automatically create two default scenarios with diverse dose-toxicity patterns. One is a “Safe” scenario, in which all doses are safe with toxicity probabilities equal to or smaller than the target \( p_T \). The true MTDC locates at the top-right corner of the dose matrix. The other is an “Ideal” scenario, in which some dose combination are tolerable but some are overly toxic and the true MTDC locates in the middle of the dose matrix. The detailed algorithm for Default Scenarios generation is provided in §4.3.1.1.

2) Logistic Regression (Figure 4.6)

Specify the four coefficients of the logistic regression, \( \beta_0, \beta_1, \beta_2 \) and \( \beta_3 \), that represent the
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

toxicity probability at the minimum candidate doses of agents 1 and 2 in the logit scale ($\beta_0$), the toxicity effect of agent 1 ($\beta_1$), the toxicity effect of agent 2 ($\beta_2$), and the toxicity effect of the interaction between the two agents ($\beta_3$), respectively. Hover mouse over each coefficient, and a description will be displayed explaining the meaning of the coefficient. Click the “Generate” button to generate the toxicity probabilities for all dose combinations. The detailed algorithm of generating scenarios through Logistic Regression is provided in §4.3.1.2.

3) Marginals & Interaction (Figure 4.7)

Specify the marginal true toxicity probabilities of agents 1 and 2 respectively and the interaction effect between the two agents, and click the “Generate” button to generate the toxicity probabilities of all pre-defined dose combinations. The detailed algorithm of generating scenarios through Marginals & Interaction is provided in §4.3.1.3.

4) Manual Construction (Figure 4.8)

After clicking the Manual Construction tab, an empty dose matrix of two agents ($n_{dose,2} \times n_{dose,1}$) will appear. Users can manually type in the true toxicity probability for each combination. Then click “Submit” button to submit the scenario.

The generated scenarios will be displayed as a scenario list and editable (Figures 4.5-4.8). Click the “Delete” button to delete the selected scenario. A bubble chart (Figures 4.5-4.8, lower right) will also be generated to visually display the whole toxicity profile for each scenario. The size of and value inside each bubble indicate the probability of toxicity at each dose combination. The three colors of bubbles indicate their positions from the true MTDC: blue, green and red bubbles mean the dose combinations are below, equal to, and above the MTDC, respectively.
Figure 4.5: Automatically generate scenarios (Default Scenarios) in the Dual-Agents Cohort-Based Designs module.
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

Figure 4.6: Generate scenarios through Logistic Regression in the Dual-Agents Cohort-Based Designs module.
4.2. User Interface and Tutorial
4.2.2. Simulation Setup

**Step 3: Generate scenarios**

* Specify the number of dose levels \( n_{\text{dose1}} \) and \( n_{\text{dose2}} \) for agent 1 and agent 2. Then select tabs below and follow instructions there to create scenarios based on the toxicity profile matrix generated through a method indicated by the tab name. Please notice that MTDC indicates the maximum tolerated dose combination.

![Image of simulation setup interface]

- **Default Scenarios**
- **Logistic Regression**
- **Marginals & Interactions**
- **Manual Construction**

* Specify the marginal toxicity probabilities of two agents and the interaction between them. Then click the **Generate** button to add a scenario based on the generated toxicity probabilities for inputs of marginal toxicity probabilities of two agents and the interaction between them.

- True toxic prob. of agent 1:
  - True toxic prob. of dose 1: 0.01
  - True toxic prob. of dose 2: 0.1
  - True toxic prob. of dose 3: 0.2

- True toxic prob. of agent 2:
  - True toxic prob. of dose 1: 0.05
  - True toxic prob. of dose 2: 0.13
  - True toxic prob. of dose 3: 0.2

- Interaction: 0.5

**Figure 4.7:** Generate scenarios through **Marginals & interactions** in the Dual-Agents Cohort-Based Designs module.
Figure 4.8: Manually generate scenario (Manual Construction) in the Dual-Agents Cohort-Based Designs module.
4.2.2.4 Launch simulation

Once the steps 1-3 are completed, users can conduct simulated clinical trials to examine the operating characteristics of the selected designs using the selected scenarios, by clicking the “Launch Simulation” button at the bottom of Simulation Setup tab (Figures 4.5-4.8). A green “Launch Successful” message will then be displayed on the screen (Figure 4.9) to indicate that the simulation has been successfully launched. Users may click the “Proceed To Simulation Results” button in the pop-up box to track the simulation processing status and simulation results.

![Launch Successful message](image)

**Figure 4.9:** “Launch Successful” message after launching simulation in the Dual-Agents Cohort-Based Designs module.
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

4.2.3 Simulation Results

In the Simulation Results tab, users can view the simulation progress and simulation results (§4.2.3.1), restore the simulation settings if needed (§4.2.3.2), and download U-Design’s proprietary report consisting of simulation results in Word format (§4.2.3.3). Hereinafter, we use the terms “simulation results” and “operating characteristics” interchangeably.

4.2.3.1 View simulation results

In the Simulation Results tab, the Running Simulations panel exhibits the progress of ongoing simulation (Figure 4.10). The ongoing simulations are displayed in ascending order by the launch time. Click the icon “×” to delete the corresponding simulation.

![Figure 4.10: Simulation progress in the Dual-Agents Cohort-Based Designs module.](image)

Once the simulations are completed, the Running Simulations panel in Figure 4.10 will disappear, green “simulation result created” massages will appear instead and stay at the same place of the Running Simulations panel unless explicitly dismissed by clicking the icon “×” at the end of the corresponding row, and the simulation results will be automatically loaded into the Simulation History panel (Figure 4.11), with the blue bold Unread sign to indicate new results. All the previously completed simulations are also listed in the Simulation History panel. Simulation results for other modules can also be viewed under the Simulation History by dropping down the “Select a module” button (Figure 4.11). Click the button to delete the selected simulation results.
4.2. User Interface and Tutorial

4.2.3. Simulation Results

Figure 4.11: Simulation Results in the Dual-Agents Cohort-Based Designs module.

Click the button to unfold the simulation results (Figure 4.12). The design settings are firstly displayed at the top of each simulation study (Figure 4.12). Then the results of simulation are shown in two ways: plots and tables.

Figure 4.12: View the simulation results in the Dual-Agents Cohort-Based Designs module.
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

Simulation Result Plots
There are two sections in the Simulation Result Plots:

A. Line plots showing four summary statistics of the simulation results for all the designs (Figure 4.13), including Prob. of Selecting MTDC, Prob. of Toxicity, Prob. of Selecting Dose-over-MTDC, and Prob. of No Selection.

B. A table of mean and standard deviation (s.d.) for the four summary statistics (Figure 4.14).

![Simulation result plots in the Dual-Agents Cohort-Based Designs module.](image)

<table>
<thead>
<tr>
<th>Summary of Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Prob. of Selecting MTDC</td>
</tr>
<tr>
<td>Prob. of Toxicity</td>
</tr>
<tr>
<td>Prob. of Selecting Dose-over-MTDC</td>
</tr>
<tr>
<td>Prob. of No Selection</td>
</tr>
</tbody>
</table>

* Mean ± Standard deviation. The statistics are calculated based on the current scenario and design setting.

![Simulation summary in the Dual-Agents Cohort-Based Designs module.](image)

Your trial designs anywhere, anytime.
A. Line plots:

- The four summary statistics are part of operating characteristics of the designs. They are explained in full detail next.

  - **Prob. of Selecting MTDC**: The probability of selecting the true MTDC, defined as the proportion of simulated trials that correctly select the true MTDC. The higher the value, the better the design.
    
    * For CI3+3 & BLRM-2d designs, the true MTDCs are defined as the dose combination levels of which the true toxicity probabilities fall into the equivalence interval \([p_T - \epsilon_1, p_T + \epsilon_2]\); if none of the dose combinations have a toxicity probability that falls into the equivalence interval, the true MTDC is defined as the dose combination with the highest toxicity probability below \(p_T\). For the PIPE design, the true MTDCs are defined as the dose combination levels with the highest toxicity probabilities lower than or equal to \(p_T\).
    
    * To compare the operating characteristics of multiple designs submitted in a simulation study, the definition of MTDC should be unified. If any of CI3+3 & BLRM-2d designs are used in the simulation, the CI3+3 and BLRM-2d might use different EI’s \([p_T - \epsilon_1, p_T + \epsilon_2]\). Then the MTDCs are defined as the dose combination levels of which the true toxicity probabilities fall into the widest equivalence interval \([p_T - \max\{\epsilon_1\}, p_T + \max\{\epsilon_2\}]\). Here, \(\max\{\cdot\}\) is taken over the designs. If none of the dose combinations fall in, the dose combination with the highest toxicity probability that is below \(p_T\) is the true MTDC. For example, consider a case in which users compare three designs, CI3+3, BLRM-2d and PIPE, in a simulation study targeting \(p_T = 0.3\). Suppose \(\epsilon_1 = 0.02\) and \(\epsilon_2 = 0.05\) for CI3+3, and \(\epsilon_1 = 0.05\) and \(\epsilon_2 = 0.03\) for BLRM-2d. In this case, the true MTDC is the dose combination levels with toxicity probabilities in \([0.3 - 0.05, 0.3 + 0.05]\); if none of the dose combinations have a toxicity probability in \([0.3 - 0.05, 0.3 + 0.05]\), the dose combination with the highest toxicity probability lower than 0.3 is the true MTDC.
    
    * For the designs that choose multiple dose combinations as the MTDCs at the end of the trial (PIPE & CI3+3), Prob. of Selecting MTDC is the percentage of simulated trials that correctly select at least one true MTDC.
    
    * If a scenario does not have any MTDC (e.g., all dose combinations have toxicity probabilities higher than the target \(p_T\)), no selection is the right decision. In this...
case, the probability of selecting the true MTDC is the probability of no selection.

- **Prob. of Toxicity**: The proportion of patients who have experienced DLT across all the simulated trials. The lower the number, the fewer patients having DLTs under the design.

- **Prob. of Selecting Does-over-MTDC**: The probability of selecting the dose combination levels above the true MTDC, defined as the percentage of simulated trials that select any dose combinations with true toxicity probabilities higher than $p_T$ at the end of the trial. The lower the value, the better the safety of the design.

- **Prob. of No Selection**: The proportion of the simulated trials in which none of the dose combination levels are selected as the MTDC. If a scenario does not have any MTDC, this values is treated as the probability of selecting the true MTDC, i.e., the correct decision.

- For each line plot, the x-axis is the index of scenario and the y-axis is the value of summary statistics. Lines with different colors represent different designs.

- The plots are interactive for better visualization.

  - Hover the mouse on a dot and a box will display the value of each design at the corresponding scenario (top left plot in Figure 4.13: Prob. of Selecting MTDC)
  
  - Hover the mouse on the design label to highlight the corresponding line and fade the others (bottom right plot in Figure 4.13: Prob. of No Selection).
  
  - Click the design label to hide the corresponding line and click again to change it back (top right plot in Figure 4.13: Prob. of Toxicity).

B. Simulation summary table: Figure 4.14 shows the mean ± sd of the summary statistics across all scenarios for each design.
### Simulation Result Tables

Full simulation results are presented mainly in tabular format arranged by scenarios (Figure 4.15), each with five sections (a bubble plot and four tables). The first section is a bubble plot that summarizes the scenario setting with dose levels of two agents and their true toxicity probabilities at each dose combination level. The middle three sections (the first three tables), from top to bottom, report the selection probability, the average number of patients treated, and the average number of toxicities (i.e. DLTs) at each dose combination, respectively. In these four sections, the green, blue and red bubbles (cells) represents doses that are the true MTDC(s), below and above the true MTDC(s), respectively. The last section reports the four trial-specific summary statistics, which are the same as those shown in the **Simulation Result Plots**, mainly from two aspects: MTDC selection and trial toxicity.

The first three tables following the bubble plot (Figure 4.15) present three summary statistics from the simulation.

- **Selection Prob.**: The proportion of simulated trials that select each dose level as the MTDC,
- **Average # of Patients Treated**: The average number of patients treated at each dose level,
- **Average # of Toxicities**: The average number of patients experienced DLT at each dose level.

The last table reports the following summary statistics for the simulation (Figure 4.15).

- **MTDC Selection**
  - **Prob. of Selecting MTDC**: The proportion of simulated trials that select the true MTDC at the end of the trial.
  - **Prob. of Selecting Does-over-MTDC**: The proportion of simulated trials that select the doses higher than the true MTDC at the end of the trial.
  - **Prob. of No Selection**: The proportion of simulated trials in which none of the dose levels are selected as the MTDC.

For detailed descriptions, please refer to **Simulation Result Plots** section above.

- **Trial Toxicity**
  - **Prob. of Toxicity**: The proportion of patients experiencing DLT across all the simulated trial. For detailed descriptions, please refer to **Simulation Result Plots** section above.
Figure 4.15: Simulation result tables in the Dual-Agents Cohort-Based Designs module.
4.2. User Interface and Tutorial
4.2.3. Simulation Results

4.2.3.2  Restore simulation

Users can restore the simulation setting from the simulation results by clicking the button at the upper right corner of each simulation results panel (yellow arrow in Figure 4.16), which will switch the display to the Simulation Setup page with the simulation settings restored. This is useful to restore existing simulation settings for reproducible results.

![Simulation Setup](image)

**Figure 4.16:** Restore simulation setup and download simulation results in the Dual-Agents Cohort-Based Designs module.

4.2.3.3  Download simulation results

There is a button at the upper right corner of each simulation results panel (green arrow in Figure 4.16). Click it to download a zip file, which includes a Word file and four line plots of summary statistic shown in Figure 4.13. The Word file is the U-Design’s proprietary report simulation report with complete simulation results under the designs and scenarios users added in the Simulation Setup page. Users could update and revise the simulation settings and results tailored for their trials or contact us for consulting services via email admin@laiyaconsulting.com.
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

4.3 Statistical Methods Review

4.3.1 Methods for Scenario Generation

In the Dual-Agents Cohort-Based Designs module, U-Design provides four methods to generate scenarios (different dose-toxicity response patterns) for simulation studies: 1) Default Scenarios based on a logistic regression, 2) Scenarios through Logistic Regression, 3) Scenarios through Marginals & Interactions, and 4) Scenarios through Manual Construction. This section describes the detailed methods of the first three methods in details.

Notation

Consider a trial combining $I$ ($I \geq 2$) dose levels of agent A, denoted by \{\(d_{A,1}, d_{A,2}, \ldots, d_{A,I}\)\}, and $J$ ($J \geq 2$) dose levels of agent B, denoted by \{\(d_{B,1}, d_{B,2}, \ldots, d_{B,J}\)\}, for dose finding. Let $d_{ij} = (d_{A,i}, d_{B,j})$ represent the combination of dose levels $i$ and $j$ for agents A and B respectively and $\pi_{ij}$ represent its true toxicity probability, for $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$.

4.3.1.1 Method for Generation of Default Scenarios

In this method, the doses of agents A and B are standardized to be in the interval $[0, 1]$, via $u_i = \frac{d_{A,i} - d_{A,1}}{d_{A,I} - d_{A,1}}$ and $v_j = \frac{d_{B,j} - d_{B,1}}{d_{B,J} - d_{B,1}}$, respectively. Therefore, the lowest dose combination is $(u_1, v_1) = (0, 0)$ and the highest dose combination is $(u_I, v_J) = (1, 1)$. We model the drug combination-toxicity relationship $\pi_{ij}$ using a four-parameter logistic model:

$$\text{logit}(\pi_{ij}) = \log(\frac{\pi_{ij}}{1 - \pi_{ij}}) = \beta_0 + \beta_1 u_i + \beta_2 v_j + \beta_3 u_i v_j,$$

(4.1)

where $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$ are four unknown parameters that represent the logit of the toxicity probability at the minimum available doses corresponding to $u_1 = v_1 = 0$ ($\beta_0$), the toxicity effect of agent A ($\beta_1$), the toxicity effect of agent B ($\beta_2$), and the toxicity effect of the interaction between two agents ($\beta_3$), respectively. Denote $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ the vector of four unknown parameters in model (4.1).

To specify the unknown values ($\beta_0, \beta_1, \beta_2, \beta_3$), we follow a procedure as follows. Firstly, we elicit with physicians four “anchor” probabilities $\pi_{1J}^*, \pi_{11}^*, \pi_{11}^*$, and $\pi_{11}^*$, corresponding to the toxicity probabilities of the four dose combinations at $(d_{A,I} = 1, d_{B,J} = 1)$, $(d_{A,I} = 1, d_{B,1} = 0)$,
4.3. Statistical Methods Review

4.3.1. Methods for Scenario Generation

\(d_{A,1} = 0, d_{B,J} = 1\), and \(d_{A,1} = 0, d_{B,1} = 0\). Under (4.1), this means

\[
\begin{align*}
\beta_0 + \beta_1 + \beta_2 + \beta_3 &= \logit(\pi_{IJ}^*) \\
\beta_0 + \beta_2 &= \logit(\pi_{11}^*) \\
\beta_0 + \beta_1 &= \logit(\pi_{I1}^*) \\
\beta_0 &= \logit(\pi_{11}^*)
\end{align*}
\]  

which can be rewritten in matrix format:

\[
A\beta = \Pi,
\]

where

\[
A = \begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0
\end{pmatrix}, \quad \beta = \begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3
\end{pmatrix}, \quad \Pi = \begin{pmatrix}
\logit(\pi_{IJ}^*) \\
\logit(\pi_{11}^*) \\
\logit(\pi_{11}^*) \\
\logit(\pi_{11}^*)
\end{pmatrix}.
\]

Then the solution of \(\beta\) can be easily solved by

\[
\hat{\beta} = A^{-1}\Pi, \quad \text{i.e.,} \quad \begin{pmatrix}
\hat{\beta}_0 \\
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\hat{\beta}_3
\end{pmatrix} = \begin{pmatrix}
\logit(\pi_{11}^*) \\
\logit(\pi_{11}^*) - \logit(\pi_{11}^*) \\
\logit(\pi_{11}^*) - \logit(\pi_{11}^*) \\
\logit(\pi_{11}^*) - \logit(\pi_{11}^*) - \logit(\pi_{11}^*) + \logit(\pi_{11}^*)
\end{pmatrix}.
\]  

In U-Design, we assume that the four “anchor” probabilities may take two default choices:

1) \(\pi_{IJ}^* = p_T, \pi_{I1}^* = \frac{p_T \times J}{I+J-1}, \pi_{11}^* = \frac{p_T \times I}{I+J-1}\) and \(\pi_{11}^* = \frac{p_T}{I+J-1}\), in which the MTD is the highest dose combination of the dose matrix; or

2) \(\pi_{IJ}^* = p_T + \frac{(1-p_T)(t+J-t-m+1)}{t+J-1-m+1}, \pi_{I1}^* = \frac{p_T \times J}{I+J-t}, \pi_{11}^* = \frac{p_T \times I}{I+J-t-1}\) and \(\pi_{11}^* = \frac{p_T \times J}{I+J-t}\), in which the MTD is in the middle of the dose matrix. Here, \(t = \frac{J}{2}\), if \(I\) is even; otherwise, \(t = \frac{I+1}{2}\).

Similarly, \(m = \frac{J}{2}\), if \(J\) is even; otherwise, \(m = \frac{J+1}{2}\).

Substitute the estimated \(\hat{\beta}\) into equation (4.1) to obtain the probability of toxicity for each dose combinations \(\pi_{ij}^*\), for \(i = 1, 2, \ldots, I\) and \(j = 1, 2, \ldots, J\). This produces two Default Scenarios.

### 4.3.1.2 Logistic Regression

Using the logistic regression (4.1), users can generate more scenarios by specifying the four parameters \(\beta = (\beta_0, \beta_1, \beta_2, \beta_3)\). Following §4.3.1.1, one can elicit the “anchor” probabilities to generate scenarios.
4.3.1.3 Marginal & Interactions

In this method, we model the dose-toxicity relationship through marginal toxicity probabilities of each agent when they are used alone and an interaction effect between the two agents.

We start by introducing some additional notation. Let $\pi_{A,i}$ and $\pi_{B,j}$ be two single-agent probabilities of DLT ascribed to $i$-th level of agent A and $j$-th level of agent B, respectively, for $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$. In the special case of no interaction (independence), the single-agent toxicities fully determine the toxicity of combinations. For dose combination $(d_{A,i}, d_{B,j})$, the probability of no DLT is $(1 - \pi_{A,i})(1 - \pi_{B,j})$. Under independence, let $\pi_{0ij}$ be the probability of no DLT under the combination $(d_{A,i}, d_{B,j})$ when the two drugs are independent; it is true that

$$\pi_{0ij} = 1 - (1 - \pi_{A,i})(1 - \pi_{B,j}) = \pi_{A,i} + \pi_{B,j} - \pi_{A,i}\pi_{B,j}. $$

On the odds scale this is equivalent to

$$ odds_{0ij} = odds_{A,i} + odds_{B,j} + odds_{A,i} \times odds_{B,j}, $$

where $odds_{0ij} = \pi_{0ij} / (1 - \pi_{0ij})$, etc. To allow interaction, one assumes

$$ odds_{ij} = odds_{0ij} \times g(\eta, d_{A,i}, d_{B,j}). $$

In U-Design, we use the same interaction $g(\cdot)$ for all dose combinations, i.e., $g(\eta, d_{A,i}, d_{B,j}) = \exp(\eta)$. Different values of $\eta$ represent different relationship between the two agents. Specifically,

- $\eta = 0$: No interaction.
- $\eta < 0$: Protective, i.e., the drug combination produces a toxic effect less than that if the drugs act independently in the body.
- $\eta > 0$: Synergistic, the drug combination produces a toxic effect greater than that if the drugs act independently in the body.

Lastly, we have toxicity probabilities for all dose combinations through $\pi_{ij} = \frac{odds_{ij}}{1 + odds_{ij}}$.

4.3.1.4 Manual Construction

We also allow users to manually input scenarios (toxicity probabilities for all dose combinations, $\pi_{ij}$). See detailed procedure in §4.2.2.3.
The product of independent beta probabilities dose escalation (PIPE) design is a Bayesian dose finding method for a combination therapy with two active agents, introduced in Mander and Sweeting (2015). The PIPE design aims to target a MTD contour such that the probabilities of toxicity for all dose combinations on this contour equal the prespecified target toxicity level $p_T$. The dose finding decision process is based on the estimated contour, and multiple dose combinations can be recommended to take forward to phase II.

4.3.2.1 Probability Model

Let $d_{A,i}$ denote the $i$-th dose level of agent A and $d_{B,j}$ denote the $j$-th dose level of agent B, $i = 1, 2, \ldots, I (I \geq 2)$ and $j = 1, 2, \ldots, J (J \geq 2)$. Assume $d_{A,i} < d_{A,i+1}$ and $d_{A,j} < d_{A,j+1}$. Let $d_{ij} = (d_{A,i}, d_{B,j})$ represent the combination of dose levels $i$ and $j$ for agents A and B respectively, and $\pi_{ij}$ represent its true toxicity probability. The toxicity is assumed to be monotonic increasing with increasing dose. That is, $\pi_{ij} \leq \pi_{i+1,j}, i = 1, 2, \ldots, I - 1, \forall j$ and $\pi_{ij} \leq \pi_{i,j+1}, j = 1, 2, \ldots, J - 1, \forall i$.

PIPE assumes $\pi_{ij}$ follows an independent beta distribution, i.e., $\pi_{ij} | a_{ij}, b_{ij} \sim \text{beta}(a_{ij}, b_{ij})$, $\forall i, j$. Here, $(a_{ij} + b_{ij})$ represents a measure of the amount of information contained in the prior, equivalent to the number of patients observed at dose $d_{ij}$ before the trial begins; and $a_{ij} / (a_{ij} + b_{ij})$ and $b_{ij} / (a_{ij} + b_{ij})$ represent the expected prior proportions of DLTs and non-DLTs at dose $d_{ij}$, respectively. In U-Design, we use a strong prior $a_{ij} = b_{ij} = 0.5$, $\forall i, j$. The reason we call $\text{beta}(0.5, 0.5)$ a strong prior is because we follow the terminology in the PIPE paper (Mander and Sweeting, 2015). Specifically, the authors use the word “strong” to contrast the weak prior in their method which corresponds to $\sum_{ij} (a_{ij} + b_{ij}) = 1$. Request to allow other priors via emailing admin@laiyaconsulting.com.

Patients are recruited into the trial sequentially in cohorts of a pre-specified size with each cohort assigned a dose combination chosen by the design. Suppose after the first $m$ cohorts, $y_{ij}^{(m)}$ patients out of $n_{ij}^{(m)}$ patients have experienced DLT for dose combination $d_{ij}$; the data up to the end of the $m$-th cohort are defined by $Data^{(m)} = \{y_{ij}^{(m)}, n_{ij}^{(m)}, i = 1, \ldots, I, j = 1, \ldots, J\}$. Then because of conjugacy and prior independence of the $\pi_{ij}$, the posterior distribution of $\pi_{ij}$ is also a beta distribution given by

$$\pi_{ij} | Data^{(m)}, a_{ij}, b_{ij} \sim \text{beta}(a_{ij} + y_{ij}^{(m)}, b_{ij} + n_{ij}^{(m)} - y_{ij}^{(m)}).$$

Your trial designs anywhere, anytime.
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

4.3.2.2 Maximum Tolerated Contour (MTC)

The PIPE design aims to locate the MTC$_pT$ corresponding to the pre-specified target probability of toxicity $p_T$, and uses MTC$_pT$ to recommend the dose level for the next cohorts. The MTC$_pT$ is defined as the boundary in the two-dimensional dose combination space that partition the space into doses with toxicity probabilities above $p_T$ or below $p_T$. The estimated MTC$_pT$ under PIPE is constrained to follow the monotonicity assumption.

For a discrete set of dose combinations, there are a finite number of locations that a contour can partition the space. And due to monotonicity assumption, only contours that satisfy monotonicity (such contours will be called the “monotonic contours”) will be considered. In general, for an $I \times J$ matrix, there are $\binom{I+J}{I}$ monotonic contours in total. For example, consider a situation where each agent has two dose levels of experimentation. There are only six possible monotonic contour choices for the MTC$_pT$, as shown in Figure 4.17. Each contour is represented by a binary matrix indicating whether doses are above the contour (1) or below (0). Define the set of all monotonic contours as $\mathcal{C}$. And let the binary matrices that are members of the set $\mathcal{C}$ be $C_s$, where $s = 1, \ldots, \binom{I+J}{I}$.

To find the most likely contour for the MTC$_pT$, consider the posterior probability that the toxicity probability is less than or equal to $p_T$ for any dose combination $d_{ij}$:

$$p_{ij}^{(m)} = \text{Prob} \left( \pi_{ij} \leq p_T \mid y_{ij}^{(m)}, n_{ij}^{(m)}, a_{ij}, b_{ij} \right).$$
4.3. Statistical Methods Review

4.3.2. The Product of Independent Beta Probabilities Dose Escalation (PIPE)

Hence, the probability that the $MTC_{py}$ is the contour defined by matrix $C_s$,

$$\alpha_s^{(m)} = \mathbb{P}(MTC = C_s | Data^{(m)}) = \prod_{i,j} \left\{ \begin{array}{ll} 1 - p_{ij}^{(m)} & \text{if } C_s[i,j] \vspace{1em} \\ p_{ij}^{(m)} & \text{if } 1 - C_s[i,j], \end{array} \right. 1 - c_s[i,j], \quad s = 1, 2, \ldots, \left( \begin{array}{c} I + J \\ I \end{array} \right), \quad (4.5)$$

where $C_s[i,j]$ is the 0-1 indicator for dose combination $d_{ij}$ in the binary matrix as shown in Figure 4.17. The underlying rationale behind the PIPE method is that dose-escalation decisions are based on the most likely $C_s$ based on $\alpha_s^{(m)}$. In other words, PIPE decides the dose finding based on the contour

$$C^{*(m)} = \arg\max_{C_s \in \mathcal{C}} \alpha_s^{(m)}. \quad (4.6)$$

4.3.2.3 Dose Finding Rules

PIPE uses $C^{*(m)}$ as the basis to guide dose finding and to choose from a set of dose combinations that are close to $C^{*(m)}$. Such set is called the admissible dose set, denoted by $\Omega^{(m)}$. In PIPE, two dose strategies are provided to define $\Omega^{(m)}$: the closest strategy and the adjacent strategy. Let $\Omega_{\text{closest}}^{(m)}$ and $\Omega_{\text{adjacent}}^{(m)}$ be the two corresponding admissible dose sets, respectively. Here, a dose combination $d_{i'j'}$ is considered closest to $C^{*(m)}$, if any of the following eight conditions is met,

a1) if $d_{i'j'}$ is above the $C^{*(m)}$, i.e., $C_s[i', j'] = 1$, and
   i. if $1 < i' \leq I, 1 < j' \leq J$, the dose combinations that are one dose level lower than $d_{i'j'}$ for only agent A or B ($d_{i'-1,j'}$ and $d_{i', j'-1}$) are below the $C^{*(m)}$, i.e., $C_s[i' - 1, j'] = C_s[i', j' - 1] = 0$; or
   ii. if $i' = 1, 1 < j' \leq J$, the dose combination that is one dose level lower than $d_{i'j'}$ for agent B ($d_{i', j'-1}$) is below the $C^{*(m)}$, i.e., $C_s[i', j' - 1] = 0$; or
   iii. if $1 < i' \leq I, j' = 1$, the dose combination that is one dose level lower than $d_{i'j'}$ for agent A ($d_{i'-1,j'}$) is below the $C^{*(m)}$, i.e., $C_s[i' - 1, j'] = 0$; or
   iv. if $d_{i'j'}$ is the lowest dose combination, i.e., $i' = j' = 1$;

a2) if $d_{i'j'}$ is below the $C^{*(m)}$, i.e., $C_s[i', j'] = 0$, and
   i. if $1 \leq i' < I, 1 \leq j' < J$, the dose combinations that are one dose level higher than $d_{i'j'}$ for only agent A or B ($d_{i'+1,j'}$ and $d_{i', j'+1}$) are above the $C^{*(m)}$, i.e., $C_s[i' + 1, j'] = C_s[i', j' + 1] = 1$; or
   ii. if $i' = I, 1 \leq j' < J$, the dose combination that is one dose level higher than $d_{i'j'}$ for agent B ($d_{i', j'+1}$) is above the $C^{*(m)}$, i.e., $C_s[i', j' + 1] = 1$; or
iii. if $1 \leq i' < I, j' = J$, the dose combination that is one dose level higher than $d_{i'j'}$ for agent A ($d_{i' + 1,j'}$) is above the $C^*_m$, i.e., $C_s[i' + 1, j'] = 1$; or
iv. if $d_{i'j'}$ is the highest dose combination, i.e., $i' = I$ and $j' = J$.

Similar, a dose combination $d_{i'j'}$ is considered adjacent to $C^*_m$, if any of the following four conditions is met,

b1) if $d_{i'j'}$ is above the $C^*_m$, i.e., $C_s[i', j'] = 1$, and
i. if $1 < i' \leq I, 1 < j' \leq J$, among the dose combinations that are a maximum of one dose level lower than $d_{i'j'}$ for both agents A and B, $d_{i-1,j}, d_{i,j-1}$ and $d_{i-1,j-1}$, there exits at least one dose combination located below the $C^*_m$, i.e., $C_s[i' - 1, j'] = 0$, $C_s[i', j' - 1] = 0$ or $C_s[i' - 1, j' - 1] = 0$; or
ii. if the dose level of agent A or B is the lowest, i.e., $i' = 1$ or $j' = 1$;

b2) if $d_{i'j'}$ is below the $C^*_m$, i.e., $C_s[i', j'] = 0$, and
i. if $1 < i' \leq I, 1 \leq j' < J$, among the dose combinations that are a maximum of one dose level higher than $d_{i'j'}$ for both agents A and B, $d_{i+1,j}, d_{i+1,j+1}$ and $d_{i+1,j+1}$, there exits at least one dose combination located above the $C^*_m$, i.e., $C_s[i' + 1, j'] = 1$, $C_s[i', j' + 1] = 1$ or $C_s[i' + 1, j' + 1] = 1$; or
ii. if the dose level of agent A or B is the highest, i.e., $i' = I$ or $j' = J$;

Figure 4.18 shows an example for two agents, each with six doses, where the solid line is $C^*_m$, the sign X’s denote the dose combination that are closest to $C^*_m$ and +’s denote the dose combinations that are adjacent but not closest to $C^*_m$. Due to the toxicity monotonicity assumption, all closest doses are adjacent.
In the PIPE paper, Mander and Sweeting (2015) provide two ways to choose one of the admissible dose combinations as the dose for the next cohort,

1) Select the next dose combination to be the admissible dose with the smallest current sample size, where sample size here is defined as both the prior and trial sample size combined, that is, $s_{ij}^{(m)} = n_{ij}^{(m)} + a_{ij} + b_{ij}$. Mathematically, this means to select the dose for the next cohort

$$d_{i^*j^*} = \underset{d_{ij} \in \Omega^{(m)}}{\text{argmin}} S_{ij}^{(m)}.$$ 

If multiple doses are returned by this function, then the dose combination administered is selected randomly from this set with equal probabilities.

2) Select the next dose combination based on a weighted randomization, where the selection of the admissible doses is weighted by the inverse of their sample size, that is,

$$\mathbb{P}\left(\text{cohort } m + 1 \text{ is allocated at } d_{ij} \mid d_{ij} \in \Omega^{(m)} \right) = \frac{S_{ij}^{-1}(m)}{\sum_{d_{ij} \in \Omega^{(m)}} S_{ij}^{-1}(m)}.$$ 

In U-Design, we take the closest dose strategy 1) to define the admissible dose set, i.e., $\Omega^{(m)} = \Omega^{(m)}_{closest}$, and choose the admissible dose with the smallest current sample size, i.e., strategy 1) above. Request to apply other dose-escalation rules via email admin@laiyaconsulting.com.

### 4.3.2.4 Dose Skipping and Safety Rules

In phase I dose-finding trials, dose skipping through the pre-defined levels of agents A and B is often prohibited. Such constraints are accommodated within the PIPE design. In U-Design, we apply the **Neighborhood Constraint**, which forces the admissible doses for the next cohort to come from a restricted set of doses that are a maximum of one dose level higher or lower than the current experimented dose both for agents A and B. Besides, U-Design does not allow diagonal escalation, i.e., escalation from $d_{ij}$ to $d_{i+1,j+1}$ is not allowed. Therefore, the admissible doses can be identified given the adjusted neighborhood constraint, and as an example, are shown in Figure 4.19 for a trial that has its current cohort doses at either (a) $d_{11}$ or (b) $d_{33}$. In example (a), the dashed box indicates the admissible doses under the current adjusted neighborhood constraint; i.e., doses $d_{12}$ and $d_{21}$; however, neither is adjacent or closest to the estimated MTD, $C^{*}(m)$. In this case, PIPE will randomly select one of those two dose combinations to be the next administered dose. In example (b), there are now three dose combinations that are closest, $d_{24}$, $d_{34}$ and $d_{43}$, and six adjacent, $d_{24}$, $d_{23}$, $d_{33}$, $d_{34}$, $d_{43}$ and $d_{42}$, that could be chosen under the adjacent strategy. Request to apply other...
Figure 4.19: The sets of admissible doses that are closest and adjacent (X), and adjacent but not closest (+) and largest (*) to $C^*(m)$ under a neighborhood constraint without diagonal escalation applied in U-Design. The dashed line shows the current neighborhood constraint (i.e. only dose combinations within the dashed box are admissible).

constraints, such as the Non-neighborhood Constraint mentioned in Mander and Sweeting (2015), via emailing admin@laiyaconsulting.com.

Additionally, a Safety Constraint is imposed to avoid any potential over-dosing. Consider the expected probability of dose combination $d_{ij}$ being above the MTC$_{p_T}$, averaged over the distribution of the monotonic contours. Denote this probability as $q_{ij}^{(m)}$ after $m$ cohorts, which is written as

$$q_{ij}^{(m)} = \sum_{C_s \in \mathcal{C}} C_s[i, j] \mathbb{P}(\text{MTC} = C_s \mid \text{Data}^{(m)}).$$

The safety constraint excludes dose combination $d_{ij}$ from the admissible dose set if $q_{ij}^{(m)} > \delta$, where $\delta$ is a prespecified constant. Mander and Sweeting (2015) have found that choosing $\delta = 0.8$ gives desired operating characteristics in the simulation studies. U-Design uses $\delta = 0.8$ by default. The trial is terminated early if there are no available dose combinations that satisfy the safety constraint.

For further safety, two additional safety rules in mTPI-2 and i3+3 are also applied in U-Design.

- [Rule 1: Dose Exclusion] If the current dose combination is considered excessively toxic, i.e., $\text{Prob}\{\pi_{ij} > p_T \mid \text{Data}^{(m)}\} > \xi$, where the threshold $\xi$ is close to 1, say 0.95, the current and all higher dose combinations $\{d_{ml} : i \leq m \leq I, j \leq l \leq J\}$ will be excluded and never be used again in the remainder of the trial.
4.3. Statistical Methods Review

4.3.2. The Product of Independent Beta Probabilities Dose Escalation (PIPE)

[Rule 2: Early Stop] If the current dose is lowest dose combination and is considered excessively toxic according to Rule 1, early stop the trial for safety.

4.3.2.5 The Recommended Phase II Doses

At the end of the trial, multiple doses can be recommended further experimentation at phase II. To do this, after the last cohort $M$ has been enrolled, $C^*(M)$ is estimated. Dose combinations that are closest from below to $C^*(M)$, have been tried during the trial and do not violate the safety constraint/rules are selected as the recommended phase II doses (RP2Ds).
4.3.3 The Bayesian Logistic Regression Method for Combination of Two Agents (BLRM-2d)

This section describes the Bayesian logistic regression method design for a combination of two active agents (BLRM-2d), proposed by Neuenschwander et al. (2015).

4.3.3.1 Probability Model

Consider a trial combining \( I(I \geq 2) \) dose levels of agent A, denoted by \( \{d_{A,1}, d_{A,2}, \ldots, d_{A,I}\} \), and \( J(J \geq 2) \) dose levels of agent B, denoted by \( \{d_{B,1}, d_{B,2}, \ldots, d_{B,J}\} \), for dose finding. Let \( d_{ij} = (d_{A,i}, d_{B,j}) \) represent the combination of dose levels \( i \) and \( j \), and \( \pi_{ij} \) represent the true toxicity probability for dose combination \( (d_{A,i}, d_{B,j}) \), for \( i = 1, 2, \ldots, I \) and \( j = 1, 2, \ldots, J \). Assume \( d_{A,i} < d_{A,i+1} \) and \( d_{B,j} < d_{B,j+1} \).

The BLRM-2d assumes a logistic model between the marginal toxicity probability of each agent and the dose levels, and the toxicity of probability of the dual agent combination is constructed by the marginal toxicity probability of each agent and the interaction between them, the same as the model in §4.3.1.3. Specifically, the relationship of the marginal toxicity probability of each agent and the dose levels is given by:

\[
\begin{align*}
\text{logit}(\pi_{A,i}) &= \log(\text{odds}_{A,i}) = \log(\alpha_1) + \beta_1 \times \log(d_{A,i}/d_{A,\text{ref}}), \quad \alpha_1, \beta_1 > 0, \\
\text{logit}(\pi_{B,j}) &= \log(\text{odds}_{B,j}) = \log(\alpha_2) + \beta_2 \times \log(d_{B,j}/d_{B,\text{ref}}), \quad \alpha_2, \beta_2 > 0,
\end{align*}
\]

where \( \alpha_1, \beta_1, \alpha_2 \) and \( \beta_2 \) are the unknown parameters, \( \pi_{A,i} \) and \( \pi_{B,j} \) are the marginal toxicity probabilities ascribed to \( i \)-th level of agent A and \( j \)-th level of agent B respectively, for \( i = 1, 2, \ldots, I \) and \( j = 1, 2, \ldots, J \), and \( d_{A,\text{ref}} \) and \( d_{B,\text{ref}} \) are the reference doses for agents A and B, respectively. U-Design uses the (ceiling of \((I+1)/2\))-th and (ceiling of \((J+1)/2\))-th level of agents A and B as default reference doses, respectively. This release users from the burden of setting reference doses manually on U-Design; however, we provide service of customized input of these values upon users requests by emailing us admin@laiyaconsulting.com. In the special case of no interaction, \( \alpha_1, \beta_1, \alpha_2, \) and \( \beta_2 \) fully determine the toxicity probability for a dose combination. For dose combination \( (d_{A,i}, d_{B,j}) \) the probability of having no DLT is \( (1 - \pi_{A,i})(1 - \pi_{B,j}) \). Hence, the probability of DLT under no interaction is

\[
\pi_{ij}^0 = 1 - (1 - \pi_{A,i})(1 - \pi_{B,j}) = \pi_{A,i} + \pi_{B,j} - \pi_{A,i} \pi_{B,j}.
\]
4.3. Statistical Methods Review

4.3.3. The Bayesian Logistic Regression Method for Combination of Two Agents (BLRM-2d)

On the odds scale, we have

\[
\text{odds}_{ij}^0 = \frac{\pi_{ij}^0}{1 - \pi_{ij}^0} = \frac{\pi_{A,i} + \pi_{B,j} - \pi_{A,i}\pi_{B,j}}{1 - (\pi_{A,i} + \pi_{B,j} - \pi_{A,i}\pi_{B,j})} \\
= \frac{\pi_{A,i} + \pi_{B,j} - \pi_{A,i}\pi_{B,j}}{(1 - \pi_{A,i})(1 - \pi_{B,j})} \\
= \frac{\pi_{A,i}}{1 - \pi_{A,i}} + \frac{\pi_{B,j}}{1 - \pi_{B,j}} + \frac{\pi_{A,i}}{1 - \pi_{A,i}} \times \frac{\pi_{B,j}}{1 - \pi_{B,j}} \\
= \text{odds}_{A,i} + \text{odds}_{B,j} + \text{odds}_{A,i} \times \text{odds}_{B,j}
\]

Adding an interaction parameter \( \eta \) has the interpretation of an odds-multiplier as follows:

\[
\text{odds}_{ij} = \text{odds}_{ij}^0 \cdot \exp(\eta).
\]

Hence, the probability of DLT at dose combination \((d_{A,i}, d_{B,j})\) is given by

\[
\pi_{ij} = \text{odds}_{ij}/(1 + \text{odds}_{ij})
\]

4.3.3.2 Likelihood and Prior Specification

Let \( n_{ij} \) and \( y_{ij} \) be the number of patients treated at dose combination \((d_{A,i}, d_{B,j})\) and the corresponding number of patients with DLTs, respectively. For observed data, \( Data \equiv \{y_{ij}, n_{ij} : i = 1, 2, \ldots, I, j = 1, 2, \ldots, J\} \), the likelihood function is the product of the binomial densities, i.e.,

\[
L(Data | \theta_1, \theta_2, \eta) = \prod_{i} \prod_{j} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{n_{ij} - y_{ij}},
\]

where \( \theta_1 = (\alpha_1, \beta_1) \) and \( \theta_2 = (\alpha_2, \beta_2) \) are vectors of unknown parameters in equations (4.7a) and (4.7b), respectively.

For the prior specification of parameters, \( \alpha_k \) and \( \beta_k \) (\( k = A \) or \( B \), denoting different agents) follow a multivariate log-normal prior, \( \pi(\theta_1) \) or \( \pi(\theta_2) \), given by

\[
\left( \log(\alpha_k) \quad \log(\beta_k) \right) \sim \text{MVN} \left( \begin{pmatrix} \mu_{k,1} \\ \mu_{k,2} \end{pmatrix}, \Sigma \right), \quad \text{where} \quad \Sigma = \begin{pmatrix} \sigma_{k,1}^2 & \rho_k \sigma_{k,1} \sigma_{k,2} \\ \rho_k \sigma_{k,1} \sigma_{k,2} & \sigma_{k,2}^2 \end{pmatrix}, \quad (4.8)
\]

where “MVN” stands for a multivariate normal distribution. The interaction parameter \( \eta \) follows a normal distribution as follows \( \eta \sim N(\mu_\eta, \sigma_\eta^2) \). In U-Design, we use the quantile-based non-informative prior calculator proposed by Neuenschwander et al. (2008) to specify the hyperparameters \((\mu_{k,1}, \mu_{k,2}, \sigma_{k,1}, \sigma_{k,2}, \rho_k)\) in (4.8) for each agent, as described in their Appendix A.1.
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

The hyperparameter calculation process is based on a set of quantiles for the probabilities of toxicity that are derived from minimally informative unimodal beta distributions. Here, a beta distribution \( X \sim \text{beta}(a,b) \) is defined as a minimally informative unimodal distribution, given a prespecified quantile \( q(p) \) of the prior distribution, if (i) \( \text{Prob}\{X < q(p)\} = p \), (ii) \( a \geq 1 \) or \( b \geq 1 \) (or both), and (iii) \( a+b \) minimal. For a given prior quantile \( q(p) \), the parameters and the quantiles of a minimally informative unimodal beta distribution can be easily obtained. If \( q(p) > p \), \( \text{beta}(a,1) \) is minimally informative unimodal if \( a = \ln(p)/\ln(q(p)) \). Alternatively, if \( q(p) < p \), \( \text{beta}(1,b) \) is minimally informative unimodal if \( b = \ln(1-p)/\ln(1-q(p)) \). Specifically, the following steps are used for this prior distribution specification for each agent, using agent A as an example:

1. Obtain the set of prior quantiles \( Q \) for the distribution of \( p_d \). In U-Design, we summarize prior information at a given dose using the median, 2.5%-th and 97.5%-th percentiles, denoted by \( q_d = \{q_d(2.5\%), q_d(50\%), q_d(97.5\%)\} \).
   (a) For the lowest dose \( d = 1 \), the prior probability of exceeding a certain threshold \( q_1(\phi_1) \) is \( \phi_1 \). In U-Design, the following default values will be used: \( \text{Prob}\{p_1 > 0.4\} = 5\% \), i.e. for the lowest dose the probability of excessive toxicity will be set to be 5 percent.
   (b) For the highest dose \( d = D \), the prior probability of falling below a certain threshold \( q_D(\phi_2) \) is \( \phi_2 \). In U-Design, the following default values will be used: \( \text{Prob}\{p_D \leq 0.2\} = 0.05 \), i.e. for the highest dose the probability of under-dosing will be set to be 5 percent.
   (c) Assuming a minimally informative unimodal beta distribution in (a) and (b) leads to prior medians for the probabilities of toxicity \( p_1 \) and \( p_D \), say \( \mu_1 = q_1(50\%) \) and \( \mu_D = q_D(50\%) \).
   (d) Prior medians \( \mu_1, \ldots, \mu_D \) are assumed to be linear in log-dose on the logit scale. This decides the minimally informative unimodal beta distributions for each dose \( d \).
   (e) For each dose \( d \), two quantiles (2.5% and 97.5%) is derived using minimally informative unimodal beta distributions with prior medians equal to \( \mu_d \).
   (f) Therefore, a set of \( D \times 3 \) quantiles are obtained, denoted by \( Q = \{q_{dk}\} \) with \( q_{dk} = q_d(\pi_k) \), \( d = 1,2,\ldots,D \), \( k = 1,2,3 \), where \( \pi_1 = 2.5\% \), \( \pi_2 = 50\% \) and \( \pi_3 = 97.5\% \).

2. For the two-parameter logistic model the above constructed quantiles \( Q \) are then compared with the quantiles \( Q' \) coming from the bivariate normal prior distribution. We will minimize the following criteria:

\[
C(Q, Q') = \max_{d,k} |q_{dk} - q'_{dk}|, \quad d = 1, 2, \ldots, D, \quad k = 1, 2, 3.
\]
4.3. Statistical Methods Review

4.3.3. The Bayesian Logistic Regression Method for Combination of Two Agents (BLRM-2d)

The minimization of $C(Q, Q')$ leads to the optimal parameter for the prior distribution $\eta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, which can be achieved by a stochastic optimization using a Metropolis algorithm (Robert and Casella, 2013).

Therefore, the posterior distribution of $(\theta_1, \theta_2, \eta)$ is given by

$$p(\theta_1, \theta_2, \eta \mid Data) \propto L(Data \mid \theta_1, \theta_2, \eta)\pi(\theta_1)\pi(\theta_2)\pi(\eta)$$

$$= \prod_{i,j} (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{n_{ij} - y_{ij}} \pi(\theta_1)\pi(\theta_2)\pi(\eta),$$

where $\pi(\theta_1)$, $\pi(\theta_2)$ and $\pi(\eta)$ are the prior distributions specified above. Using Markov chain Monte Carlo (MCMC) simulation, the posterior samples could be drawn for $\theta_1, \theta_2, \eta$ and posterior inference can be made based on the samples.

4.3.3.3 Dose Finding Rules

Suppose the target probability of DLT is $p_T$, BLRM-2d divides the probability interval $(0, 1)$ into three categories: under-dosing $p_{ij} \in (0, p_T - \epsilon_1]$, target toxicity $p_{ij} \in (p_T - \epsilon_1, p_T + \epsilon_2]$, excessive and unacceptable toxicity $p_{ij} \in (p_T + \epsilon_2, 1)$. After each patient cohort is enrolled and toxicity data are observed, the next dose will be selected depending on the Targeted Toxicity Maximization Subject to Escalation with Overdose Control (EWOC). That is, select the dose for the next cohort patients as the one that maximizes the posterior probability of falling into the targeted interval, i.e.,

$$\arg\max_{i,j} \text{Prob}\{\pi_{ij} \in (p_T - \epsilon_1, p_T + \epsilon_2] \mid Data\}$$

subject to the constraint that the probability of overdosing (i.e., excessive and unacceptable toxicity) does not exceed a predefined threshold $p_{EWOC}$, i.e.,

$$\text{Prob}\{\pi_{ij} \in (p_T + \epsilon_2, 1) \mid Data\} \leq p_{EWOC}.$$ Here, $\text{Prob}\{\cdot\}$ is calculated based on posterior distribution of $(\theta_1, \theta_2, \eta)$.

4.3.3.4 Skipping and Safety Rules

In phase I dose-finding trials, dose skipping and diagonal escalation are often prohibited. To this end, U-Design defines the admissible doses for the next cohort as a set of doses that are at most one dose level higher or lower than the current dose for both agents A and B. In addition, U-Design does not allow diagonal escalation. See Figure 4.20 for an illustration. In example (a), the current dose combination is $d_{11}$ and the admissible doses are $d_{11}, d_{12}$ and $d_{21}$; in example (b), the current dose is $d_{33}$ and the admissible doses are $d_{34}, d_{43}, d_{33}, d_{24}, d_{42}, d_{23}, d_{32}$ and $d_{22}$, a total of eight doses. The trial is terminated early if there are no available doses in the admissible dose set or no doses in the admissible set satisfy the EWOC constraint.

Your trial designs anywhere, anytime.
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

Figure 4.20: The set of admissible doses. Small dots (·) denote the pre-defined dose combinations for the trial, a large dot (●) denotes the current dose, and squares and the large dot (□ and ●) denote the admissible doses for the next cohort patients.

For further safety, two additional safety rules in mTPi-2 and i3+3 are also applied in U-Design.

– [Rule 1: Dose Exclusion] If the current dose combination is considered excessively toxic, i.e., \( \text{Prob}\{\pi_{ij} > p_T \mid Data^{(m)}\} > \xi \), where the threshold \( \xi \) is close to 1, say 0.95, the current and all higher dose combinations \( \{d_{ml} : i \leq m \leq I, j \leq l \leq J\} \) are excluded and never used again in the remainder of the trial.

– [Rule 2: Early Stop] If the current dose is lowest dose combination and is considered excessively toxic according to Rule 1, early stop the trial for safety.

In Rules 1 and 2, \( \text{Prob}\{\pi_{ij} > p_T \mid Data\} \) is a function of the cumulative distribution of \( \text{beta}(\alpha_0 + y_{ij}, \beta_0 + n_{ij} - y_{ij}) \). In U-Design, \( \alpha_0 = \beta_0 = 1 \) is used.

4.3.3.5 The MTDC Selection

At the end of the trial, the dose combination \( d_{i^*j^*} = (d_{A,i^*}, d_{B,j^*}) \) is selected as the MTDC if it maximizes the posterior probability of toxicity rate falling into the targeted interval, i.e., \( d_{i^*j^*} = \arg\max_{i,j} \text{Prob}\{\pi_{ij} \in (p_T - \epsilon_1, p_T + \epsilon_2) \mid Data\} \) among all doses that have been used and do not violate the EWOC rule.
4.3.4 The Combo i3+3 Design (CI3+3)

The CI3+3 design is a rule-based design for finding the maximum tolerated dose combination (MTDC) for dual-agents dose-finding trials, co-developed by Laiya Consulting Inc. It adopts the dose-escalation rules of i3+3 (Liu et al., 2020) and extends them from one dimension to two dimensions.

4.3.4.1 Review of i3+3 Design

We first give a brief review of the i3+3 decision rules (Liu et al., 2020), upon which the CI3+3 design is anchored. The i3+3 design defines an equivalence interval $\text{EI} = [p_T - \epsilon_1, p_T + \epsilon_2]$ with the target probability of toxicity $p_T$ and two small fractions, $\epsilon_1$ and $\epsilon_2$, and allocates the next cohort of patients based on the relationship between toxicity probability observed on the current cohort of patients and the equivalence interval. Specifically, suppose dose $d$ is currently used in the trial to treat patients, and $y_d$ patients have experienced dose limiting toxicities (DLTs) out of $n_d$ patients that have been treated. Based on EI, the i3+3 design identifies the appropriate dose for the next cohort of patients according to the following five simple rules.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Decision</th>
<th>Next dose level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_d/n_d$ below EI</td>
<td>Escalation(E)</td>
<td>$d + 1$</td>
</tr>
<tr>
<td>$y_d/n_d$ inside EI</td>
<td>Stay(S)</td>
<td>$d$</td>
</tr>
<tr>
<td>$y_d/n_d$ above EI and $y_{d-1}/n_d$ below EI</td>
<td>Stay(S)</td>
<td>$d$</td>
</tr>
<tr>
<td>$y_d/n_d$ above EI and $y_{d-1}/n_d$ inside EI</td>
<td>De-escalation(D)</td>
<td>$d - 1$</td>
</tr>
<tr>
<td>$y_d/n_d$ above EI and $y_{d-1}/n_d$ above EI</td>
<td>De-escalation(D)</td>
<td>$d - 1$</td>
</tr>
</tbody>
</table>

Here, a value is below the EI means that the value is smaller than $(p_T - \epsilon_1)$, the lower bound of the EI. A value is inside the EI means that the value is larger than or equal to $(p_T - \epsilon_1)$ but smaller than or equal to $(p_T + \epsilon_2)$. A value is above the EI mean that the value is larger than $(p_T + \epsilon_2)$, the upper bound of the EI.

4.3.4.2 Design Algorithm

For a dual-agents dose-finding trial, suppose $I$ dose levels of agent A, denoted by $\{d_{A,1}, \ldots, d_{A,I}\}$, and $J$ dose levels of agent B, denoted by $\{d_{B,1}, \ldots, d_{B,J}\}$, are to be investigated. Assume $d_{A,i} < d_{A,i+1}$, and $d_{B,j} < d_{B,j+1}$. Let $d_{ij} = (d_{A,i}, d_{B,j})$ denote the combination of $i$-th dose level for agent A and $j$-th dose level for agent B, and let $\pi_{ij}$ denote its true toxicity probability, for
Module 4. Dual-Agents Dose-Finding Designs with Toxicity Endpoint and Cohort Enrollment

\[ i = 1, 2, \ldots, I \text{ and } j = 1, 2, \ldots, J. \] Generally, toxicity is assumed to be monotonic increasing with increasing dose of each agent; That is, \( \pi_{ij} \leq \pi_{i+1,j} \), \( i = 1, 2, \ldots, I - 1 \), \( \forall j \) and \( \pi_{ij} \leq \pi_{i,j+1} \), \( j = 1, 2, \ldots, J - 1 \), \( \forall i \). This results in a partial order. Suppose at any moment in the trial, dose combination \( d_{ij} = (d_{A,i}, d_{B,j}) \) is currently used to treat patients and a total of \( n_{ij} \) patients have been assigned to dose combination \( d_{ij} \). Let \( y_{ij} \) be the number of patients (among \( n_{ij} \)) with DLTs.

The CI3+3 design consists of two stages, the run-in stage and the adaptive stage. In CI3+3, patients are enrolled in cohorts, say three patients per cohort. To begin the trial, CI3+3 enrolls the first cohort patients at the starting dose combination. For simplicity, suppose the starting dose is the lowest dose combination \( d_{11} \).

**Stage I: Run-in Stage**

In Stage I, CI3+3 escalates the dose along a prespecified path in order to explore the dose-combination space quickly. Within this path, the doses are fully ordered with monotonic toxicity. Therefore, existing designs for single-agent dose-finding trials can be used, say the i3+3 design (Liu et al., 2020).

The path can be chosen based on some pre-clinical and clinical information, such as the mechanism of the two agents and the clinical conjecture of MTDC locations. See Figure 4.21 for three possible paths. When we have little information about the path in Stage I, both the blue and red paths in Figure 4.21 might be good choices, which can be written mathematically as follow:

\[
P_1 = \{d_{11} \Rightarrow d_{12} \Rightarrow \cdots \Rightarrow d_{1J} \Rightarrow d_{2J} \Rightarrow \cdots \Rightarrow d_{IJ}\},
\]

\[
P_2 = \{d_{11} \Rightarrow d_{21} \Rightarrow \cdots \Rightarrow d_{11} \Rightarrow d_{12} \Rightarrow \cdots \Rightarrow d_{IJ}\}.
\]

If a single path is chosen in stage I, CI3+3 uses the i3+3 design to conduct dose finding along the doses on the path, until

1) a “de-escalation” or a “stay” decision is suggested; or

2) the highest dose along the path is reached.

In addition, we introduce an algorithm using both the blue and red paths \( P_1 \) and \( P_2 \) as an alternative choice. In CI3+3, we use the first few cohorts of patients to explore dose combinations in both paths.
4.3. Statistical Methods Review
4.3.4. The Combo i3+3 Design (CI3+3)

and later select one of them for further escalation. The algorithm is implemented in U-Design and given below.

I(a) For the first cohort of patients treated at \( d_{11} \), suppose the toxicity outcomes are \((y_{11}, n_{11})\), where \( n_{11} \) is the number of patients treated at \( d_{11} \) and \( y_{11} \) is the corresponding number of DLTs,

- if the i3+3 design gives the “Escalation” decision, enroll two cohorts of patients and treat them at dose combinations \( d_{12} \) and \( d_{21} \) simultaneously to explore two paths \( P_1 \) and \( P_2 \) at the same time;
- Otherwise, Stage I ends and Stage II starts.

I(b) If Stage I does not ends in Step I(a), based on the toxicity outcomes at doses \( d_{12} \) and \( d_{21} \), \((y_{12}, n_{12})\) and \((y_{21}, n_{21})\), five possible decisions are considered.

- If the i3+3 design gives the “Escalation” decision for both \( d_{12} \) and \( d_{21} \), select the path with the higher observed toxicity rate \((y/n)\) and escalate according to the corresponding path. If \( d_{12} \) and \( d_{21} \) have the same rate (i.e., \( y_{12}/n_{12} = y_{21}/n_{21} \)), select one of them at random. The other escalation and path are abandoned. For example, if dose combination \( d_{12} \) is selected, escalate the dose to \( d_{13} \) along the path \( P_1 \) and stop the escalation to \( d_{21} \).

- If the i3+3 design gives the “Escalation” decision for one dose combination and “De-escalation” for another, select the path that has the dose combination with the “Escalation” decision. For example, if “Escalation” for \( d_{12} \) and “De-escalation” for \( d_{21} \), select path \( P_1 \) and assign the next cohort patients to the next higher dose level of \( P_1 \), the \( d_{13} \). The other decision and path are abandoned.

- If the i3+3 design gives the “Stay” decision for both dose combinations, end Stage I and select the dose combination \( d_{i^*j^*} \) that has the smaller difference between the observed toxicity rate and the target rate as the starting dose of Stage II, i.e., \( d_{i^*j^*} = \text{argmin}_{d_{ij} \in \{d_{12}, d_{21}\}} \left| \frac{y_{ij}}{n_{ij}} - p_T \right| \). If \( d_{12} \) and \( d_{21} \) have the same difference, select one of them randomly. Abandon the other dose combination.

- If the i3+3 design gives the “Stay” decision for one dose combination and “Escalation” or “De-escalation” for the other, end Stage I and select the dose combination with “Stay” decision as the starting dose of Stage II. For example, if “Stay” for \( d_{12} \) and “Escalation” for \( d_{21} \), starts Stage II with \( d_{12} \) as its starting dose. Abandon the other dose combination \( d_{21} \).
If the i3+3 design gives the “De-escalation” decision for both doses, de-escalate to \( d_{11} \) and end Stage I.

(c) If Stage I does not ends in Step Ib, a path must have been selected, say \( P^* \in \{ P_1, P_2 \} \). Then escalate the dose along the path \( P^* \) according to the i3+3 design. Suppose \( d_{ij} \) is currently used, end Stage I and start Stage II if either of the following two conditions is satisfied:

1) when the decision is “De-escalation” or “Stay” based on the observed toxicity outcomes \((y_{ij}, n_{ij})\) at the dose combination \( d_{ij} \) according to the i3+3 design;

2) when the highest dose combination \( d_{I,J} \) is reached.

**Stage II: Adaptive Dose-Finding Stage**

In Stage II, the path is no longer needed and dose finding is expanded to all the dose combinations. The dose-assignment decisions in CI3+3 are on the basis of an extended i3+3 decision rule using a utility function. Specifically, suppose \( d_{ij} \) is being used as the current dose. Based on the observed toxicity outcomes \((y_{ij}, n_{ij})\) at the dose combination \( d_{ij} \), the decision can be made according to the i3+3 design, “Escalation”, “Stay” or “De-escalation”. Define the “candidate” dose set of \( d_{ij} \) as the doses adjacent to it, denoted by \( \Omega_{ij} \).

\[
\Omega_{ij} = \{ i', j' : i', j' \in \mathbb{N}^+, i' \leq I, j' \leq J, |i' - i| \leq 1, |j' - j| \leq 1, -1 \leq (i' - i) + (j' - j) \leq 1 \}.
\]

If \( 1 < i < I \) and \( 1 < j < J \), it can be written in a simple way,

\[
\Omega_{ij} = \{ d_{i-1,j}, d_{i,j-1}, d_{i,j}, d_{i-1,j+1}, d_{i,j+1}, d_{i+1,j-1}, d_{i+1,j}, d_{i,j+1} \}.
\]

See Figure 4.22 for examples. In example (a), the current dose is \( d_{35} \), and its candidate dose set is \( \Omega_{35} = \{ d_{25}, d_{34}, d_{35}, d_{44}, d_{45} \} \), a total of five dose combinations. In example (b), the current dose is \( d_{33} \), and its candidate dose set is \( \Omega_{33} = \{ d_{23}, d_{32}, d_{33}, d_{42}, d_{34}, d_{43} \} \), a total of eight dose combinations. Furthermore, candidate doses can be divided into three categories corresponding to the three dose escalation decisions, respectively. They are

\[
\Omega_{ij}^E = \{ i', j' : i', j' \in \mathbb{N}^+, i' \leq I, j' \leq J, |i' - i| \leq 1, |j' - j| \leq 1, (i' - i) + (j' - j) = 1 \},
\]

\[
\Omega_{ij}^S = \{ i', j' : i', j' \in \mathbb{N}^+, i' \leq I, j' \leq J, |i' - i| \leq 1, |j' - j| \leq 1, (i' - i) + (j' - j) = 0 \},
\]

\[
\Omega_{ij}^D = \{ i', j' : i', j' \in \mathbb{N}^+, i' \leq I, j' \leq J, |i' - i| \leq 1, |j' - j| \leq 1, (i' - i) + (j' - j) = -1 \},
\]

where \( \Omega_{ij}^E, \Omega_{ij}^S \) and \( \Omega_{ij}^D \) are three candidate dose sets corresponding to “Escalation”, “Stay” and “De-escalation” decisions, respectively. In simplicity, if \( 1 < i < I \) and \( 1 < j < J \), these three sets are.
Figure 4.22: The set of candidate doses. In this case, two agents are to be tested, each with five dose levels. The dashed box indicates the candidate doses, denoted by the star (*) or large dot (●).

These candidate doses are classified into three categories highlighted with three different colors: green stands for candidate doses of “Escalation”, blue for “Stay”, and orange for “De-escalation”.

\[ \Omega_{E}^{ij} = \{d_{i+1,j}, d_{i+1,j+1}\}, \Omega_{S}^{ij} = \{d_{ij}, d_{i-1,j+1}, d_{i+1,j-1}\}, \Omega_{D}^{ij} = \{d_{i-1,j}, d_{i,j-1}\} \]

respectively. In Figure 4.22 (a), \( \Omega_{E}^{35} = \{d_{45}\} \), \( \Omega_{S}^{35} = \{d_{35}, d_{44}\} \), and \( \Omega_{D}^{35} = \{d_{25}, d_{34}\} \); in (b), \( \Omega_{E}^{33} = \{d_{34}, d_{43}\} \), \( \Omega_{S}^{33} = \{d_{32}, d_{24}, d_{42}\} \), and \( \Omega_{D}^{33} = \{d_{32}, d_{23}\} \). Once we obtain the i3+3 decision \( A \in \{E, S, D\} \) based on the toxicity outcomes (\( y_{ij} \) out of \( n_{ij} \) patients have DLTs) observed at the current dose \( d_{ij} \), the dose combinations in the candidate dose set \( \Omega_{ij}^{A} \) are all candidate doses for the next cohort.

For safety, a **Dose Exclusion Rule** is applied in CI3+3. Specifically, if the current dose combination is considered excessive toxic, i.e., \( \text{Prob}\{\pi_{ij} > p_T \mid \text{Data}\} > \xi \), where the threshold \( \xi \) is close to 1, say 0.95, the current and all higher dose combinations \( \{d_{ml} : i \leq m \leq I, j \leq l \leq J\} \) are excluded (from the candidate dose set) and never be used again in the remainder of the trial. Hereinafter, \( \text{Prob}\{\pi_{ij} > p_T \mid \text{Data}\} \) is a function of the cumulative distribution of \( \text{Beta}(\alpha_0 + y_{ij}, \beta_0 + n_{ij} - y_{ij}) \). In U-Design, \( \alpha_0 = \beta_0 = 1 \) is used.

CI3+3 assigns the next cohort of patients in Stage II as follows. Let \( A \in \{E, S, D\} \) denote the decision of the i3+3 design for the current dose.

II(a) If there are two or more dose combinations in \( \Omega_{ij}^{A} \), select the dose combination with the
highest utility. In U-Design, we use the utility function,

\[ U(y_{ij}, n_{ij}, p_T) = \begin{cases} 
1, & \text{if } n_{ij} = 0; \\
1/n_{ij}, & \text{if } n_{ij} > 0.
\end{cases} \]

In this way, CI3+3 prefers to assign patients to dose combinations that have been least tested, for better exploration of the dose-combination space. If two or more dose combinations are tied with the same utility, select any one randomly with equal probabilities.

II(b) If only one dose combination is available in \( \Omega^A_{ij} \), i.e., \( \Omega^A_{ij} = \{d_{ij}\} \), select that dose combination \( d_{ij} \) for the treatment of the next cohort patients.

- In the special case of \( A = S \) and \( \Omega^A_{ij} = \{d_{ij}\} \), and two cohorts of patients have already been treated at \( d_{ij} \), in order to avoid getting trapped at the dose \( d_{ij} \) as much as possible, an additional Annealing Rule is applied, see next.

II(c) If no dose combinations are left in \( \Omega^A_{ij} \), i.e., \( \Omega^A_{ij} = \emptyset \),

- If \( A = D \) and \( \Omega^D_{ij} = \emptyset \), modify the decision to “Stay” \( (A = S) \) because there are no available lower dose combinations to de-escalate to. And follow the steps IIa-IIb to assign the next cohort of patients.
- If \( A = E \) and \( \Omega^E_{ij} = \emptyset \), modify the decision to “Stay” \( (A = S) \) because there are no available higher dose combinations to escalate to. And follow the steps IIa-IIb to assign the next cohort of patients.

Annealing Rule

An additional Annealing Rule is applied to avoid dose assignment getting to be trapped at a single dose. In brief, the rule allows dose assignment to leave \( d_{ij} \) for other dose combinations with certain probabilities. This rule is only applied with \( A = S \) and \( \Omega^S_{ij} = \{d_{ij}\} \). That is, stay and the only candidate dose is \( d_{ij} \). In CI3+3, we treat “S” as a decision to either stay at the same dose, or go to a dose for which the toxicity in unordered (could be higher or lower) relative to \( d_{ij} \). For example, \( d_{i+1,j-1} \) and \( d_{i-1,j+1} \) are im-ordered to \( d_{ij} \). But if \( \Omega^S_{ij} = \{d_{ij}\} \), it means these two doses have been excluded. To this end, we consider an annealing algorithm that consider doses that are un-ordered but further away from \( d_{ij} \) to encourage exploration.

i) Specify \( \Omega^S_{ij} = \Omega^S_{ij} = \{d_{ij}\} \);

ii) Sequentially check doses from \( \{d_{i+1,k}, k = j-2, \ldots, 1\} \), until a dose is available (not excluded) and add it into \( \Omega^S_{ij} \) as new candidate dose combination for the next cohort of patients;

iii) Sequentially check doses from \( \{d_{k,j+1}, k = i-2, \ldots, 1\} \), until a dose is available (not
4.3.4. The Combo i3+3 Design (CI3+3)

excluded) and add it into $\Omega_{ij}^{S0}$ as new candidate dose combination for the next cohort of patients;

iv) Go to II(a) to continue dose finding.

**Trial Termination**

For safety, an *Early Stop Rule* is applied in CI3+3. If the current dose is the lowest dose combination and considered excessively toxic, i.e., $\text{Prob}\{\pi_{11} > p_T \mid Data\} > \xi$, where the threshold $\xi$ is close to 1, say 0.95, stop the trial early. Similarly, $\text{Prob}\{\pi_{11} > p_T \mid Data\}$ is a function of the cumulative distribution of $\text{Beta}(\alpha_0 + y_{11}, \beta_0 + n_{11} - y_{11})$. In U-Design, $\alpha_0 = \beta_0 = 1$ is used.

The trial also stops when the pre-specified maximum sample size $N$ is reached.

**4.3.4.3 MTDC Selection**

At the end of the trial, multiple doses can be recommended as the MTDCs. CI3+3 uses the same MTDC selection method as that of PIPE in §4.3.2 at the end of the trial.
Part II

Phase Ib Expansion Cohort Designs
5. Multiple Cohort Expansion

5.1 Introduction

In modern early-phase clinical trials, often times multiple doses of a new drug are tested in multiple indications to identify the promising doses and arms for phase II or phase III trials. Traditionally, each dose or indication is tested separately in a single trial, resulting in multiple protocols and multiple trials. This module describes a new solution, the multiple cohort expansion (MUCE) design (Lyu et al., 2020). MUCE is a Bayesian solution for cohort expansion trials or the master protocol trials, in which multiple dose(s) and multiple indication(s) are expanded in parallel. It’s built on Bayesian hierarchical models with multiplicity control to adaptively borrow information across patient groups from different indications treated with different dose to achieve three major goals:

1. Control the type I error rate (probability of selecting an unpromising drug for further development);
2. Increase the power (probability of selecting a promising drug for further development);
3. Reduce sample size.

As a comprehensive statistical solution, MUCE can be used to calculate the sample size or power, and to conduct interim and final data analyses for making critical decisions. These can be applied in any clinical trials with two or more arms, including:

1. Phase Ib trials with multiple expansion cohorts;
2. Phase II trials with multiple arms;
3. Master protocols including basket, umbrella, and platform trials;
5.2 User Interface and Tutorial

5.2.1 Overview

Entering the MUltiple Cohort Expansion page, users will see four main tabs: 1) Introduction, 2) Case Study, 3) Quick Demo and 4) Data Analysis (Figure 5.1). In the Introduction tab, a general description of MUCE design, its application and benefits is provided. Then three real-world trials that used MUCE as their trial designs are listed in the Case Study tab, to demonstrate the superiority of MUCE when compared with other designs (§5.2.2). Next, in the Quick Demo tab, a demo of the sample size calculation function of MUCE is given, which is based on a simple numerical search algorithm (§5.2.3). Last, in the Data Analysis tab (§5.2.4), users could estimate response rates and corresponding posterior probabilities and perform Bayesian hypothesis testing, to conduct interim and final analyses for critical decision-making, such as selecting optimal treatment arm(s).

![Figure 5.1: Overview of the Multiple Cohort Expansion module.](image)
Module 5. Multiple Cohort Expansion

5.2.2 Case Study

The Case Study tab lists three real-world cases that apply MUCE (Figure 5.2). In each case study, MUCE is demonstrated to have superior operating characteristics in terms of reducing sample size and controlling the type I error rate (probability of selecting an unpromising drug for further development). Click “Learn More” button in each case box to open and download a PDF file with the detailed descriptions of the case study.

![Figure 5.2: Three real world case studies in the Multiple Cohort Expansion module.](image-url)
5.2. User Interface and Tutorial

5.2.3. Quick Demo

This is a demo of the sample size calculation function of the MUCE module on U-Design. In this demo, all dose-indication arms are assumed to have the same reference response rates and target response rates, therefore all arms should have the same sample size, if the type I error rates and powers are also prespecified the same across all arms. It is a simplified situation and upon these assumption, the sample size of each arm can be easily found through a numerical search algorithm, such as the binary search algorithm. In this quick demo, only limited values are allowed for some input parameters. All limits will be removed in the full version of MUCE module.

Setup  Select the number of doses \(n_{dose}\) and the number of indications \(n_{ind}\) from dropdown boxes, resulting in a total of \(n_{dose} \times n_{ind}\) dose-indication arms for MUCE designs. Then specify the reference response rate (historical control rate) for each indication \(R_{ref}\) and the target response rate for each arm \(R_{target}\). Hover mouse over each trial parameter, and a description will be displayed explaining the meaning of the parameter (Figure 5.3). The detailed explanation of the above four input arguments and their limited values allowed to be selected are provided in Table 5.1. Upon selection, click the “Submit” button to calculate the sample size for each arm using MUCE design to reach the desired type I error rate \(\alpha\) and power (Figure 5.3).

![Figure 5.3: Set trial parameters in the Quick Demo of the Multiple Cohort Expansion module.](image)
Table 5.1: Input trial parameters in the **Quick Demo** of the **Multiple Cohort Expansion** module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{sim}$</td>
<td>Number of simulated trials</td>
<td>The number of simulated trials to be conducted for each scenario.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In this quick demo, it is fixed at 1,000.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Type I error rate</td>
<td>The probability of rejecting null when the null hypothesis is true.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In this quick demo, it is fixed at 0.05.</td>
</tr>
<tr>
<td>$power$</td>
<td>Power</td>
<td>$power = 1 - \beta$, where $\beta$ is the type II error rate, i.e., the probability of rejecting null when the alternative hypothesis is true. In this quick demo, it is fixed at 0.8.</td>
</tr>
<tr>
<td>$n_{dose}$</td>
<td>Number of doses</td>
<td>The number of doses evaluated in the trial. Two values are available for selection In this quick demo, $n_{dose} \in {1, 2}$.</td>
</tr>
<tr>
<td>$n_{ind}$</td>
<td>Number of indications</td>
<td>The number of indications expanded in the trial. Two values are available for selection In this quick demo, $n_{ind} \in {2, 3}$.</td>
</tr>
<tr>
<td>$R_{ref}$</td>
<td>Reference response rate</td>
<td>The reference response rate (also called the historical control rate) is the largest rate considered to be not promising. Three values are available for selection In this quick demo, $R_{ref} \in {0.1, 0.2, 0.3}$.</td>
</tr>
<tr>
<td>$R_{target}$</td>
<td>Target response rate ($R_{target} &gt; R_{ref}$)</td>
<td>The target response rate is the smallest rate considered to be promising. Three values are available for selection in this version, In this quick demo, $R_{target} \in {0.2, 0.3, 0.4}$.</td>
</tr>
</tbody>
</table>

**Results**  The results are displayed in two parts (Figure 5.4):

1. Sample size of MUCE and its comparison with that of Simon’s two-stage design.
   - First line lists the values of seven trial parameters in Table 5.1 specified above.
   - A table gives the sample size suggested for MUCE design, to reach the desired type I error and power, using the Simon’s two-stage design as benchmark.
   - A description of sample size justification in protocol language.

2. Sample size searching process based on the binary search algorithm.
   - A table lists all the sample size that have been tried in an ascending order, and their corresponding calculated type I error rates and powers.
   - The minimum sample size that reaches the desired type I error rate and power is selected and highlighted in orange background.
5.2. User Interface and Tutorial

5.2.3. Quick Demo

Summary of Performance

\[ n_{\text{min}} = 1000, \quad \alpha = 0.05, \quad \text{power} = 0.8, \quad n_{\text{sample}} = \frac{2}{n}, \quad R_{\text{ref}} = 0.1, \quad R_{\text{target}} = 0.2. \]

Comparison of sample sizes required for Simon's 2-Stage and MUCE

<table>
<thead>
<tr>
<th>Design</th>
<th>Sample size</th>
<th>Type I error</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simon's 2-stage</td>
<td>36 36 36 36</td>
<td>0.04 0.03 0.03 0.02</td>
<td>0.941 0.941 0.941 0.941</td>
</tr>
<tr>
<td>MUCE</td>
<td>36 36 36 36</td>
<td>0.04 0.03 0.03 0.02</td>
<td>0.941 0.941 0.941 0.941</td>
</tr>
</tbody>
</table>

* The index for each arm \( x \) represents the cohort in the \( m \)th indication with \( m \)th dose level.

The above table shows MUCE can save up to 51.2% sample size, compared to Simon's 2-stage design.

- The MUCE design is used to calculate the sample size for the cohort expansion trial, in which 2 doses and 8 indications, a total of 6 arms, are expanded in parallel. The MUCE design requires a sample size of 40 for each arm to maintain a type I error rate of 0.05 and power of 0.8, given a reference response rate of 0.1 and target response rate of 0.2. Compared to Simon's 2-stage design, which needs 82 patients, the MUCE design can save up to 51.2% sample size.

Note:
1. The sample size of Simon's two-stage design is the expected sample size, which incorporates the futility stopping in the first stage. It could be calculated by \( N_1 \times \text{PET} + N_1 \times (1 - \text{PET}) \), where \( N_1 \) is the sample size of first stage, \( N \) is the total sample size, and \( \text{PET} \) is the probability of early termination due to futility.
2. The global null scenario is the case where all arms are not promising with the response rates equal to 0.1.
3. The global alternative scenario is the case where all arms are promising with the response rates equal to 0.2.
4. We select the minimum sample size required to maintain an averaged power of 0.8 and type I error of 0.05 for each arm.

Below table illustrates the process of searching the optimal sample size for MUCE

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Type I error</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>36 36 36 36</td>
<td>0.04 0.03 0.02</td>
<td>0.941 0.941 0.941</td>
</tr>
</tbody>
</table>

* The row with a background color is recommended by MUCE.

Figure 5.4: MUCE sample size calculation results in the Quick Demo of the Multiple Cohort Expansion module.
Module 5. Multiple Cohort Expansion

5.2.4 Data Analysis

The data analysis includes response rates estimation, Bayesian hypothesis tests, and optimal selection of treatment arms, for interim or final analyses, all based on the MUCE design.

**Setup** In the Step 1, select numbers of doses and indications \( n_{dose} \in \{1, 2, 3, 4, 5\} \) and \( n_{ind} \in \{1, 2, \ldots, 20\} \) from the drop-down boxes, respectively. Upon selection, an input table of the observation data will be automatically generated below the Step 2. And users could manually type in the reference response rate \( (R_{ref}) \) for each indication, and the observed numbers of responses and patients for each dose-indication arm. The label-name of each arm can also be changed by users’ needs. See 5.5 for illustration. Click the “Submit” button to launch the analysis.

![Data Analysis Setup](image)

**Figure 5.5:** Set parameters in the Data Analysis of the Multiple Cohort Expansion module.
5.2. User Interface and Tutorial

5.2.4. Data Analysis

**Results** The analysis results are displayed in tables (Figure 5.6).

- The first three columns demonstrate the label-name, and the indexes of dose level and indication, for each arm, respectively.
- The next two columns demonstrate the inputted reference response rate \( (R_{\text{ref}}) \), the observed numbers of responds and patients \( (r/n) \), for each arm, respectively. Also, the response rate of each arm is calculated in ratio.
- The last four columns demonstrate the data analysis results based on the MUCE design, including
  - \( P_{H_1} \): Posterior probability of the alternative hypothesis that the true response rate is larger than the reference response rate. If \( P_{H_1} \) is large enough, such as \( P_{H_1} > 0.95 \), this arm is selected for further investigation (The arm with orange background color); otherwise, it is not selected.
  - \( P_{\text{mean}} \): Estimated posterior mean of response rate for each arm.
  - \( P_{\text{lower}} \) and \( P_{\text{upper}} \): The lower and upper boundaries of the interval of the response rate for each arm based on MUCE.

<table>
<thead>
<tr>
<th>Arm</th>
<th>( I_{\text{dose}} )</th>
<th>( I_{\text{indication}} )</th>
<th>( R_{\text{ref}} )</th>
<th>( r/n ) (ratio)</th>
<th>( P_{H_1} )</th>
<th>( P_{\text{mean}} )</th>
<th>( P_{\text{lower}} )</th>
<th>( P_{\text{upper}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.3</td>
<td>3 / 4 = 0.75</td>
<td>0.96</td>
<td>0.69</td>
<td>0.31</td>
<td>0.99</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0.2</td>
<td>1 / 4 = 0.25</td>
<td>0.67</td>
<td>0.29</td>
<td>0.01</td>
<td>0.62</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0.3</td>
<td>2 / 4 = 0.5</td>
<td>0.81</td>
<td>0.48</td>
<td>0.11</td>
<td>0.85</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0.2</td>
<td>1 / 6 = 0.17</td>
<td>0.56</td>
<td>0.22</td>
<td>0</td>
<td>0.48</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>0.25</td>
<td>1 / 6 = 0.17</td>
<td>0.49</td>
<td>0.24</td>
<td>0.01</td>
<td>0.51</td>
</tr>
</tbody>
</table>

**Note:**
- If \( P_{H_1} \) is large enough, such as \( P_{H_1} > 0.95 \), this arm is selected for future trial considerations (The arm with background color); otherwise, it is not selected.
- Hover the mouse over the table header to see the description of each column.
  - \( \text{Arm} \): Name of each arm
  - \( I_{\text{dose}} \): The index of dose level
  - \( I_{\text{indication}} \): The index of indication
  - \( R_{\text{ref}} \): The reference response rate for each indication
  - \( r/n \) (ratio): Number of responses / Number of patients (Response rate)
  - \( P_{H_1} \): posterior probability of the alternative hypothesis that the true response rate is larger than the reference rate
  - \( P_{\text{mean}} \): Estimated posterior mean of response rate for each arm
  - \( P_{\text{lower}} \): The lower bound of the credible interval of the response rate for each arm based on MUCE
  - \( P_{\text{upper}} \): The upper bound of the credible interval of the response rate for each arm based on MUCE

**Figure 5.6:** Results in the **Data Analysis** of the **Multiple Cohort Expansion** module.

*Your trial designs anywhere, anytime.*

207
5.3 Statistical Methods Review

5.3.1 Multiple Cohort Expansion (MUCE) Method

The multiple cohort expansion (MUCE) (Lyu et al., 2020) approach was proposed as a design or analysis method for phase 1b multiple expansion cohort trials, which investigate one or more doses of a new investigational drug in patients from with different indications (cancer types and/or biomarker status). The MUCE design is based on a class of Bayesian hierarchical models that adaptively borrow information across different dose-indication arms. Statistical inference is directly based on the posterior probability of each arm being efficacious, facilitating the decision making that decides which arm to select for further testing.

5.3.1.1 Probability Model

Consider a phase Ib trial that evaluates \( J \) different dose levels of a new drug in \( I \) different indications. Let \((i, j)\) denote the cohort arm for indication \( i \) and dose level \( j \), \( i = 1, \ldots, I, \; j = 1, \ldots, J \). The total number of arms is \( K = I \times J \). Suppose \( n_{ij} \) patients have been treated in arm \((i, j)\), and \( y_{ij} \) of them are responders. Let \( p_{ij} \) denote the true and unknown response rate for the arm \((i, j)\). We assume \( y_{ij} \) follows a binomial distribution conditional on \( n_{ij} \) and \( p_{ij} \),

\[
y_{ij} \mid n_{ij}, p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij})
\]

Whether dose level \( j \) is effective for indication \( i \) can be examined by the following hypothesis test:

\[
H_{0,ij} : p_{ij} \leq \pi_{i0} \quad \text{versus} \quad H_{1,ij} : p_{ij} > \pi_{i0} , \quad (5.1)
\]

where \( \pi_{i0} \) is the reference response rate for indication \( i \).

We perform the hypothesis test (11.6) under a formal Bayesian testing framework. Let \( \lambda_{ij} \) be a binary and random indicator of the hypothesis, such that \( \lambda_{ij} = 0 \) (or 1) represents that hypothesis \( H_{0,ij} \) (or \( H_{1,ij} \)) is true. Firstly, a prior model for \( p_{ij} \) is built under each hypothesis. Let \( \theta_{ij} = \log \left( \frac{p_{ij}}{1-p_{ij}} \right) \) denote the log-odds of the response rate. The null hypothesis \( p_{ij} \leq \pi_{i0} \) is equivalent to \( \theta_{ij} \leq \theta_{i0} \), and the alternative hypothesis is equivalent to \( \theta_{ij} > \theta_{i0} \), where \( \theta_{i0} = \log \left( \frac{\pi_{i0}}{1-\pi_{i0}} \right) \).

Conditional on \( \lambda_{ij} \), MUCE assume

\[
\theta_{ij} \mid \lambda_{ij} = 0 \sim \text{Trunc-Cauchy}(\theta_{i0}, \gamma; (-\infty, \theta_{i0}]), \\
\theta_{ij} \mid \lambda_{ij} = 1 \sim \text{Trunc-Cauchy}(\theta_{i0}, \gamma; (\theta_{i0}, \infty)),
\]

where \( \text{Trunc-Cauchy}(\theta, \gamma; A) \) denotes a Cauchy distribution with location \( \theta \) and scale \( \gamma \) truncated to interval \( A \).
5.3. Statistical Methods Review

5.3.1. Multiple Cohort Expansion (MUCE) Method

Secondly, prior models for the probabilities of the hypotheses, \( \Pr(\lambda_{ij} = 1) \) and \( \Pr(\lambda_{ij} = 0) \), are constructed. To borrow strength across dose levels and indications, we construct a hierarchical prior model for \( \lambda_{ij} \). A natural and conventional Bayesian approach is to impose a common prior for the probability of \( \{\lambda_{ij} = 1\} \), which shrinks the probabilities to a common value. To better exploit the data structure in multiple expansion cohort trials, we propose to differentiate the borrowing strength from two factors: dose and indication. To achieve this, we use a probit model for \( \lambda_{ij} \). Let \( Z_{ij} \) be a latent random variable, and \( \lambda_{ij} = I(Z_{ij} < 0) \), where \( I(\cdot) \) is an indicator function. We model

\[
Z_{ij} \sim N(\xi_i + \eta_j, \sigma_0^2).
\]

Here, \( E(Z_{ij}) = \xi_i + \eta_j \), in which \( \xi_i \) characterizes the effect of indication \( i \) and \( \eta_j \) of dose \( j \). The indication-specific effects and dose-specific effects are then separately modeled by common priors,

\[
\xi_i \mid \xi_0, \sigma_\xi \sim \text{N}(\xi_0, \sigma_\xi^2), \quad \text{and} \quad \eta_j \mid \eta_0, \sigma_\eta \sim \text{N}(\eta_0, \sigma_\eta^2).
\]

Lastly, we put hyperpriors on \( \xi_0 \) and \( \eta_0 \), \( \xi_0 \sim N(\mu_{\xi_0}, \sigma_{\xi_0}^2) \) and \( \eta_0 \sim N(\mu_{\eta_0}, \sigma_{\eta_0}^2) \).

In brief, the entire hierarchical models are summarized in the following equations:

Likelihood:
\[
y_{ij} \mid n_{ij}, p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij});
\]
Transformation:
\[
\theta_{ij} = \logit(p_{ij}), \theta_{i0} = \logit(p_{i0});
\]
Prior for (\( \theta_{ij} \mid \lambda_{ij} \)):  
\[
\theta_{ij} \mid \lambda_{ij} = 0 \sim \text{Trunc-Cauchy}(\theta_{i0}, \gamma; (-\infty, \theta_{i0}]), \quad \theta_{ij} \mid \lambda_{ij} = 1 \sim \text{Trunc-Cauchy}(\theta_{i0}, \gamma; (\theta_{i0}, \infty));
\]
Prior for \( \lambda_{ij} \):
\[
\lambda_{ij} = \begin{cases} 0, & \text{if } Z_{ij} < 0, \\ 1, & \text{if } Z_{ij} \geq 0 \end{cases} \quad (5.2)
\]

Latent probit regression:
\[
Z_{ij} \mid \xi_i, \eta_j, \sigma_\xi^2 \sim \text{N}(\xi_i + \eta_j, \sigma_0^2);
\]
Indication-specific effects:
\[
\xi_i \mid \xi_0, \sigma_\xi^2 \sim \text{N}(\xi_0, \sigma_\xi^2);
\]
Dose-specific effects:
\[
\eta_j \mid \eta_0, \sigma_\eta^2 \sim \text{N}(\eta_0, \sigma_\eta^2);
\]
Hyperpriors:
\[
\xi_0 \mid \mu_{\xi_0}, \sigma_{\xi_0}^2 \sim \text{N}(\mu_{\xi_0}, \sigma_{\xi_0}^2), \quad \eta_0 \mid \mu_{\eta_0}, \sigma_{\eta_0}^2 \sim \text{N}(\mu_{\eta_0}, \sigma_{\eta_0}^2).
\]

In U-Design, the values of the hyperparameters \( \gamma = 2.5, \mu_{\xi_0} = 0, \mu_{\eta_0} = 0, \sigma_0^2 = 1, \sigma_\xi^2 = 1, \sigma_\eta^2 = 1, \sigma_{\xi_0}^2 = 1, \sigma_{\eta_0}^2 = 1 \) and \( \sigma_{\eta_0}^2 = 1 \) are used by default.
5.3.1.2 Trial Design

Suppose \( L \geq 0 \) interim looks are planned, and the \( l \)-th interim analysis is conducted after \( n^{l}_{ij} \) patients have been enrolled in arm \( k \). Let \( D^{l} \equiv \{(n^{l}_{ij}, y^{l}_{ij}) : i = 1, 2, \ldots, I; j = 1, 2, \ldots, J\} \) denote the observed data at interim analysis \( l \), where \( y^{l}_{ij} \) is the number of responders among the \( n^{l}_{ij} \) patients. Denote \( D^{L+1} \equiv \{(n^{L+1}_{ij}, y^{L+1}_{ij}) : i = 1, 2, \ldots, I; j = 1, 2, \ldots, J\} \) the observed data at the end of the trial, where \( n^{L+1}_{ij} \) is the prespecified maximum sample size for arm \((i,j)\) and \( y^{L+1}_{ij} \) is the total number of responders. The proposed MUCE design with \( L \) interim looks is describe as follows:

1. Enroll \( n^{1}_{ij} \) patients in \((i,j)\)-th arm, \( i = 1, 2, \ldots, I; j = 1, 2, \ldots, J \).
2. Given the data \( D^{l} \) at the \( l \)-th interim look, \( l = 1, 2, \ldots, L \),
   (a) \textit{[Futility stopping]} If the posterior probability that the hypothesis of arm \((i,j)\), \( H_{1,ij} \), is true (i.e., \( \lambda_{ij} = 1 \)) is small, i.e.,
       \[
P_{H_{1}} = Pr\{\lambda_{ij} = 1 \mid D^{l}\} < P_{\text{futility}},
       \]
   stop the accrual to the \( k \)-th arm for futility;
   (b) \textit{[Efficacy stopping]} If the posterior probability that the hypothesis of arm \((i,j)\), \( H_{1,ij} \), is true (i.e., \( \lambda_{ij} = 1 \)) is large, i.e.,
       \[
P_{H_{1}} = Pr\{\lambda_{ij} = 1 \mid D^{l}\} < P_{\text{efficacy}},
       \]
   stop the accrual to the \( k \)-th arm for efficacy;
   (c) Otherwise, continue to enroll patients until reaching the next interim analysis.
3. Once the maximum sample size is reached or all the arms have stopped, evaluate the efficacy for each arm based on all the observed data. If the posterior probability that that the hypothesis of arm \( k \), \( H_{1,ij} \), is true (i.e., \( \lambda_{ij} = 1 \)) is large, i.e.,
   \[
P_{H_{1}} = Pr\{\lambda_{ij} = 1 \mid D^{L+1}\} > \phi_{ij},
   \]
   arm \((i,j)\) is declared efficacious and promising; otherwise, it is considered not promising.
Part III

Phase II Single-Arm Continuous Monitoring
6. Bayesian Efficacy Monitoring with Predictive Probability

6.1 Bayesian Efficacy Monitoring via Predictive Probability

This section describes the Bayesian Efficacy Monitoring via Predictive Probability (henceforth referred to as PP) proposed by (Lee and Liu, 2008). PP design possesses good operating characteristics. At the same time the design is more flexible compared with traditional two- or three-stage designs which can be difficult to follow exactly because the response has to be evaluated at pre-specified fixed number(s) of patients.

6.1.1 Model

Denote $p$ as the response rate. Assume $p$ follows a beta prior, $p \sim \text{Beta}(a_0, b_0)$. It represents the investigator’s previous knowledge or belief of the efficacy of the new regimen. The quantity $a_0/(a_0 + b_0)$ reflects how informative the prior is. The quantities $a_0$ and $b_0$ can be considered as the number of response and the number of nonresponses, respectively. Thus, $a_0 + b_0$ can be considered as a measure of the amount of information contained in the prior. The larger the value of $a_0 + b_0$, the more informative the prior and the stronger the belief it contains.

Let $X$ denote the number of responses among the current enrolled $n$ patients, so we have $X$ follow a binomial distribution, $X \sim \text{Binomial}(p, n)$. Consequently, the posterior distribution of response rate $p$ follows a new beta distribution,

$$p|n, X = x \sim \text{Beta}(a_0 + x, b_0 + n - x).$$  \hfill (6.1)

Set a maximum accrual of patients to $N (N \geq n)$. Thus, the number of responses ($Y$) in the
potential \( m (m = N - n) \) future patients follows a beta-binomial distribution,

\[
Y | n, m, X = x \sim \text{Beta} - \text{Binomial}(m, a_0 + x, b_0 + n - x).
\]

(6.2)

When there are \( i \) responses in the remaining \( m \) patients, i.e., when \( Y = i \), we can get the posterior distribution of response rate \( p \),

\[
p|x = x, Y = i \sim \text{Beta}(a_0 + x + i, b_0 + N - x - i).
\]

(6.3)

Let \( p_0 \) denote a reference response rate, the effect of the standard treatment. Therefore, through (6.2) and (6.3), \( PP \) can be calculated as follows:

\[
PP = \sum_{i=0}^{m} Pr(Y = i | X = x) I\{Pr(p > p_0 | X = x, Y = i) \geq \theta}\),
\]

(6.4)

where \( \theta \) is the probability threshold for declaring efficacy at the end of the trial; and \( I\{\star\} \) is the indication function, which equals to 1 if the condition satisfies; otherwise, equals to 0.

For example, in a phase II trial, an investigator plans to enroll a maximum of \( N = 15 \) patients into the study. At a given time, \( x = 2 \) responses are observed in \( n = 10 \) patients. We use the prior \( \text{Beta}(0.5, 0.5) \) and the efficacy declaration threshold \( \theta = 0.7 \). Therefore, \( PP \) of declaring efficacy (say, \( > p_0 = 0.3 \)) is 0.03, see Table 6.1 for the detail calculation process.

**Table 6.1:** Illustration of Calculating \( PP \) \((N = 15, n = 10, x = 2, p_0 = 0.3, \theta = 0.7)\)

| \( Y \) | \( X + Y \) | \( Pr(Y = i | X = x) \) | \( p|x = x, Y \sim \text{Beta}(a, b) \) in (6.3) | \( Pr(p > p_0 | X = x, Y = i) \) | Indicator\(^1\) | Prod\(^2\) |
|---|---|---|---|---|---|---|
| 0 | 2 | 0.338 | 2.5 | 13.5 | 0.071 | 0 | 0 |
| 1 | 3 | 0.338 | 3.5 | 12.5 | 0.203 | 0 | 0 |
| 2 | 4 | 0.206 | 4.5 | 11.5 | 0.404 | 0 | 0 |
| 3 | 5 | 0.088 | 5.5 | 10.5 | 0.624 | 0 | 0 |
| 4 | 6 | 0.026 | 6.5 | 9.5 | 0.804 | 1 | 0.026 |
| 5 | 7 | 0.004 | 7.5 | 8.5 | 0.917 | 1 | 0.004 |
| Total | \( 1 \) | \( 0.03 \) |

\(^1\) Indicator denotes \( I\{Pr(p > p_0 | X = x, Y = i) \geq \theta\}\).

\(^2\) Prod denotes \( Pr(Y = i | X = x) \times I\{Pr(p > p_0 | X = x, Y = i) \geq \theta\}\).

### 6.1.2 Decision Criteria

For efficacy monitoring using PP, the following two decision rules are introduced:

- **Early stopping for futility**: the trial will be stopped early and the treatment is declared inefficacious if \( PP < P_L \), where \( P_L \) is chosen as a small positive constant. \( PP < P_L \)

Your trial designs anywhere, anytime.
indicates that it is unlikely the response rate will be larger than $p_0$ at the end of the trial given the current information. When this happens, we may as well terminate the trial.

- **Early stopping for efficacy**: the trial will be stopped early and the treatment is declared efficacious if $PP > P_U$, where $P_U$ is chosen as a large positive constant. $PP > P_U$ indicates that it has a high probability to conclude that the treatment is efficacious at the end of the study, if the same trend as the current data continues. That is, the current collected data provides enough evidence to stop the trial early due to efficacy.

And the details about how to setup $P_L$, $P_T$ and $P_U$ see the future function Search.

### 6.1.3 Design

With any number of patients before the end, we can calculate a value of $PP$, and then decide whether to early stop and declare efficacy or futility by comparing $PP$ with $P_L$ and $P_U$. Exactly as the flexibility of PP design, there is not a fixed trial design. Any cohort size is adaptable. And even the cohort size can be one, so it allows continuous monitoring of the trial outcome. See the next subsection for details.

### 6.1.4 An Example

Consider a example that a study is expected to enroll 20 patients. During the trial, after 10 patients have assessed their primary endpoint, when there are new patients’ outcomes, the decision of early stop for efficacy or futility will be made by comparing the boundary values obtained based on $PP$ with the actual responses of primary endpoint.

If input parameters as shown on the left panel of Figure 6.1 and click Submit, we can get the result on the right panel of Figure 6.1. The futility and efficacy boundary values are shown in Table 6.2.

Specifically, if the number of responses is less than or equal to the futility boundary, the study may be early stopped for futility (e.g., when there are 15 patients having been assessed with less than or exactly 4 responses, early stopping for futility is permitted in this trial.); if the number of responses is more than or equal to the efficacy boundary, the study may be early stopped for efficacy (e.g., when there are 16 patients having been assessed with more than or exactly 8 responses, early stopping for efficacy is permitted in this trial.). If the trial don’t stop early for futility or efficacy, and more than or exactly 8 responses are observed in final 20 patients, the treatment will be considered effective, otherwise futile.
Module 6. Bayesian Efficacy Monitoring with Predictive Probability

**Figure 6.1:** An Example: Bayesian Efficacy Monitoring by Predictive Probability

**Table 6.2:** Futility and Efficacy Boundary Values by Predictive Probability

<table>
<thead>
<tr>
<th>Early Futility Boundary</th>
<th>Number of patients (with primary endpoint assessed)</th>
<th>10 ~ 11</th>
<th>12 ~ 14</th>
<th>15 ~ 16</th>
<th>17 ~ 18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early stop for futility, if number of responses</td>
<td>&lt;= 2</td>
<td>&lt;= 3</td>
<td>&lt;= 4</td>
<td>&lt;= 5</td>
<td>&lt;= 6</td>
</tr>
<tr>
<td>Early Efficacy Boundary</td>
<td>Number of patients (with primary endpoint assessed)</td>
<td>10 ~ 12</td>
<td>13 ~ 15</td>
<td>16 ~ 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Early stop for efficacy, if number of responses</td>
<td>&gt;= 6</td>
<td>&gt;= 7</td>
<td>&gt;= 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficacy Boundary</td>
<td>reaching the maximum sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Declaring efficacy, if number of responses</td>
<td>&gt;= 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7. Bayesian Efficacy Monitoring with Posterior Probability

7.1 Bayesian Efficacy Monitoring via Posterior Probability

This section describes the Bayesian Efficacy Monitoring via Posterior Probability (henceforth referred to as PoP). PoP design possesses good operating characteristics, more flexible compared with traditional two- or three-stage designs which can be difficult to follow exactly because the response has to be evaluated at pre-specified fixed number(s) of patients, same as PP design (see Section 6.1).

7.1.1 Model

Denote \( \theta \) as the response rate. Assume \( \theta \) follows a prior beta distribution, \( \text{Beta}(a_0, b_0) \). It represents the investigator’s previous knowledge or belief of the efficacy of the new regimen. The quantity \( a_0/(a_0 + b_0) \) reflects how informative the prior is. The quantities \( a_0 \) and \( b_0 \) can be considered as the number of response and the number of nonresponses, respectively. Thus, \( a_0 + b_0 \) can be considered as a measure of the amount of information contained in the prior. The larger the value of \( a_0 + b_0 \), the more informative the prior and the stronger the belief it contains.

Let \( X \) denote the number of responses in current \( n \) patients, so we have \( X \) follow a binomial distribution, \( X \sim \text{Binomial}(\theta, n) \). Consequently, the posterior distribution of response rate \( \theta \) follows a new beta distribution,

\[
\theta|n, X = x \sim \text{Beta}(a_0 + x, b_0 + n - x).
\]

7.1.2 Decision Criteria

For efficacy monitoring using posterior probability, the following three decision rules are introduced:

\[
\theta|n, X = x \sim \text{Beta}(a_0 + x, b_0 + n - x).
\]
Module 7. Bayesian Efficacy Monitoring with Posterior Probability

- **Early stopping for futility**: let $\theta_{\text{fut}}$ be the reference response rate for futility monitoring and $P_{\text{fut}}$ be the probability confidence threshold for futility stopping. The trial should be stopped early and the treatment is declared inefficacious if

$$Pr(\theta > \theta_{\text{fut}}|n, x) \leq P_{\text{fut}}.$$ 

- **Early stopping for efficacy**: let $\theta_{\text{eff}}$ be the reference response rate for efficacy monitoring and $P_{\text{eff}}$ be the probability confidence threshold for efficacy stopping. The trial should be stopped early and the treatment is declared efficacious if

$$Pr(\theta > \theta_{\text{eff}}|n, x) \geq P_{\text{eff}}.$$ 

- **Criterion for declaring efficacy** at the end of the trial: let $\theta_{\text{eff.final}}$ be the reference response rate and $P_{\text{eff.final}}$ be the probability confidence threshold for declaring efficacy at the end of the trial. The treatment is declared efficacious if

$$Pr(\theta > \theta_{\text{eff.final}}|n, x) \geq P_{\text{eff.final}}.$$ 

For example, assume that there is a clinical trial which has enrolled 10 patients ($n = 10$) and among these 10 patients 2 patients responds ($x = 2$). We use the prior $a_0 = 0.5, b_0 = 0.5$. So the posterior probability of $\theta$ is as follows $\theta|n = 10, X = 2 \sim Beta(2.5, 8.5)$. If we use the $\theta_{\text{fut}} = 0.3$ as the response rate for futility, so the posterior probability of response rate being higher than 0.3 is $Pr(p > 0.3|n = 10, X = 2) = 0.25$. If we use the futility threshold $P_{\text{fut}} = 0.3$, the trial will be stopped early.

### 7.1.3 Design

With any number of patients before the end, we can calculate values of

$$Pr(\theta > \theta_{\text{fut}}|n, x), Pr(\theta > \theta_{\text{eff}}|n, x) \text{ and } Pr(\theta > \theta_{\text{eff.final}}|n, x),$$

and then decide whether to early stop and declare efficacy or futility by comparing them with $P_{\text{fut}}, P_{\text{eff}}$ and $P_{\text{eff.final}}$. Exactly as the flexibility of PoP design, there is not a fixed trial design. Any cohort size is adaptable. And even the cohort size can be one, so it allows continuous monitoring of the trial outcome. See the next subsection for details.

*Your trial designs anywhere, anytime.*
7.1.4 An Example

Consider an example that a study is expected to enroll 20 patients. During the trial, after 10 patients have assessed their primary endpoint, when there are new patients’ outcomes, the decision of early stop for efficacy or futility will be made by comparing the boundary values obtained based on PoP with the actual responses of primary endpoint.

If input parameters as shown on the left panel of Figure 7.1 and click Submit, we can get the result on the right panel of Figure 7.1. The futility and efficacy boundary values are shown in Table 7.1.

Figure 7.1: An Example: Bayesian Efficacy Monitoring by Posterior Probability

Table 7.1: Futility and Efficacy Boundary Values by Posterior Probability

<table>
<thead>
<tr>
<th>Early Futility Boundary</th>
<th>Number of patients (with primary endpoint assessed)</th>
<th>10 ~ 13</th>
<th>14 ~ 16</th>
<th>17 ~ 19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early stop for futility, if number of responses</td>
<td>&lt;= 2</td>
<td>&lt;= 3</td>
<td>&lt;= 4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Early Efficacy Boundary</th>
<th>Number of patients (with primary endpoint assessed)</th>
<th>10 ~ 12</th>
<th>13 ~ 15</th>
<th>16 ~ 17</th>
<th>18 ~ 19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early stop for efficacy, if number of responses</td>
<td>&gt;= 5</td>
<td>&gt;= 6</td>
<td>&gt;= 7</td>
<td>&gt;= 8</td>
<td></td>
</tr>
</tbody>
</table>

| Efficacy Boundary reaching the maximum sample size | Declaring efficacy, if number of responses | >= 9   |
Specifically, if the number of responses is less than or equal to the futility boundary, the study may be early stopped for futility (e.g., when there are 17 patients having been assessed with less than or exactly 4 responses, early stopping for futility is permitted in this trial.); if the number of responses is more than or equal to the efficacy boundary, the study may be early stopped for efficacy (e.g., when there are 18 patients having been assessed with more than or exactly 8 responses, early stopping for efficacy is permitted in this trial.). If the trial don’t stop early for futility or efficacy, and more than or exactly 9 responses are observed in 20 patients, the treatment will be considered effective, otherwise futile.
8. Bayesian Toxicity Monitoring

8.1 Bayesian Toxicity Monitoring via Posterior Probability

This section describes the Bayesian Toxicity Monitoring via Posterior Probability. This design is mostly the same as PoP design (see Section 7.1), the only difference being that this design is used to monitor toxicity but PoP design monitors efficacy. So this design also possesses good operating characteristics, more flexible compared with traditional two- or three-stage designs which can be difficult to follow exactly because the response has to be evaluated at pre-specified fixed number(s) of patients.

8.1.1 Model

Denote $\theta$ as the toxicity rate. Assume $\theta$ follows a prior beta distribution, $\theta \sim \text{Beta}(a_0, b_0)$. It represents the investigator’s previous knowledge or belief of the toxicity of the new regimen. The quantity $a_0/(a_0 + b_0)$ reflects how informative the prior is. The quantities $a_0$ and $b_0$ can be considered as the number of DLTs and the number of non-DLTs, respectively. Thus, $a_0 + b_0$ can be considered as a measure of the amount of information contained in the prior. The larger the value of $a_0 + b_0$, the more informative the prior and the stronger the belief it contains.

Let $X$ denote the number of DLTs in current $n$ patients, $X \sim \text{Binomial}(\theta, n)$. Consequently, the toxicity distribution of toxicity rate $\theta$ follows a new beta distribution, $\theta|n, X = x \sim \text{Beta}(a_0 + x, b_0 + n - x)$.

For toxicity monitoring using toxicity probability, the trial should be stopped if $Pr(\theta > \theta_{\text{max}}|n, x) \geq \theta_T$. 

221
8.1.2 Design

With any number of patients before the end, we can calculate a value of $Pr(\theta > \theta_{\text{max}}|n, x)$ then decide whether or early stop for excessive toxicity by comparing them with $\theta_T$. Exactly as the flexibility of this design, there is not a fixed trial design. Any cohort size is adaptable. And even the cohort size can be one, so it allows continuous monitoring of the trial outcome. See the next subsection for details.

8.1.3 An Example

Consider an example that a study is expected to enroll 20 patients. During the trial, after 10 patients have assessed their primary endpoint, when there are new patients’ outcomes, the decision of early stop for efficacy or futility will be made by comparing the boundary values obtained based on PoP of toxicity with the actual DLTs of primary endpoint.

If input parameters as shown on the left panel of Figure 8.1 and click Submit, we can get the result on the right panel of Figure 8.1. The futility and efficacy boundary values are shown in Table 8.1.

**Figure 8.1:** An Example: Bayesian Toxicity Monitoring by Posterior Probability

Specifically, if the number of DLTs is more than or equal to the toxicity boundary, the study may be early stopped for excessive toxicity (e.g., when there are 14 patients having been assessed with more than or exactly 6 DLTs, early stopping for excessive toxicity is permitted in this trial.).
### Table 8.1: Futility and Efficacy Boundary Values by Posterior Probability

<table>
<thead>
<tr>
<th>Early Toxicity Boundary</th>
<th>10</th>
<th>11 ~ 13</th>
<th>14 ~ 16</th>
<th>17 ~ 19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early stop for excessive toxicity, if number of DLTs</td>
<td>$\geq 4$</td>
<td>$\geq 5$</td>
<td>$\geq 6$</td>
<td>$\geq 7$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Toxicity Boundary reaching the maximum sample size</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaring excessive toxicity, if number of DLTs</td>
<td>$\geq 8$</td>
</tr>
</tbody>
</table>
Part IV

Phase II/III Seamless Designs
9. Phase II/III Seamless Designs for Binary Outcome

Laiya extended the simulation scheme introduced in Thall and Simon (1994) for phase 2-3 seamless design, including three arms in the phase 2 stage, two doses (high and low) and placebo. At the end of the phase 2 stage, the design will make a go/no-go decision based on Bayesian go/no-go criteria first and then select one of high and low doses as the treatment arm in phase 3 based on Bayesian selection criteria.

9.1 Binary Outcome

With binary outcome, a two-sided hypothesis z-test will be performed based on data of the selected treatment dose integrated from both phase 2 and 3 and data of control arm only from phase 3 at the end of phase 3 stage,

$$H_0 : p_T = p_C \quad \text{vs} \quad H_1 : p_T \neq p_C,$$

where $p_T$ and $p_C$ represent response probabilities of the selected treatment dose and placebo respectively.

9.1.1 Model

Let $N_{H2}$, $N_{L2}$ and $N_{C2}$ denote sample sizes, $Y_{H2}$, $Y_{L2}$ and $Y_{C2}$ numbers of patients with response and $p_H$, $p_L$ and $p_C$ response probabilities for three arms in phase 2 (H, L, C represent high dose, low dose and control). The sampling models are:

$$Y_{H2} \sim Bin(N_{H2}, p_H),$$

$$Y_{L2} \sim Bin(N_{L2}, p_L),$$

226
\[ Y_{C2} \sim Bin(N_{C2}, p_C); \]

Priors in simulation:

\[ p_H \sim Beta(\alpha_{H0}, \beta_{H0}), \]
\[ p_L \sim Beta(\alpha_{L0}, \beta_{L0}), \]
\[ p_C \sim Beta(\alpha_{C0}, \beta_{C0}). \]

The conditional posterior distribution of \( p \) is (here we suppress the subscript H, L and C):

\[ p \mid N_2, Y_2 \sim Beta(\alpha_0 + Y_2, \beta_0 + N_2 - Y_2). \]

### 9.1.2 Decision Criteria

Bayesian go/no-go and selection decision with binary outcome from phase 2 that Laiya designed are as below:

- **Bayesian go/no-go criteria** at the end of phase 2 stage based on two indicators, \( h.go \) and \( l.go \). If \( h.go = 1 \) or \( l.go = 1 \), go to phase 3. Otherwise, not go.

Let \( h = Pr(p_H > p_C + \delta_0| \text{data}) \) and \( l = Pr(p_L > p_C + \delta_0| \text{data}) \).

\[ h.go = \begin{cases} 
1, \text{ if } h \geq \eta_1 \\
\sim Bin(1, h.go.p), \text{ if } \eta_2 < h < \eta_1 \\
0, \text{ if } h \leq \eta_2 
\end{cases} \]

\[ l.go = \begin{cases} 
1, \text{ if } l \geq \eta_1 \\
\sim Bin(1, l.go.p), \text{ if } \eta_2 < l < \eta_1 \\
0, \text{ if } l \leq \eta_2 
\end{cases} \]

where \( h.go.p = \frac{h - \eta_2}{\eta_1 - \eta_2}, l.go.p = \frac{l - \eta_2}{\eta_1 - \eta_2} \) and \( \delta_0 \) denotes the expected difference between the probabilities of treatment dose and placebo.

- **Bayesian selection criteria** after making go decision based on one indicator, \( h.select \). If \( h.select = 1 \), select the high dose \((T = H)\). Otherwise, select the low dose \((T = L)\).

\[ h.select = \begin{cases} 
1, \text{ if } Pr(p_H > p_L| \text{data}) > \xi \\
0, \text{ if } Pr(p_H > p_L| \text{data}) \leq \xi 
\end{cases} \]
Module 9. Phase II/III Seamless Designs for Binary Outcome

- (Criteria of z.test) Let $N_T^3$, $N_C^3$ denote sample sizes and $Y_T^3$, $Y_C^3$ numbers of patients with response for two arms in phase 3 (T, C represent selected treatment dose and control). The sampling models are the same,

$$Y_T^3 \sim Bin(N_T^3, p_T),$$
$$Y_C^3 \sim Bin(N_C^3, p_C).$$

And the estimated probabilities of two arms are,

$$\hat{p}_T = \frac{Y_T^2 + Y_T^3}{N_T^2 + N_T^3},$$
$$\hat{p}_C = \frac{Y_C^3}{N_C^3}.$$

If $1 - \Phi(Z) < z.test.\alpha/2$, we will think the selected treatment dose and placebo are significantly different, where $\Phi(*)$ denotes the standard normal distribution function and

$$Z = \frac{|\hat{p}_T - \hat{p}_C|}{\sqrt{\hat{p}_T(1 - \hat{p}_T)/(N_T^2 + N_T^3) + \hat{p}_C(1 - \hat{p}_C)/N_C^3}}.$$

9.1.3 Program Input and Output

9.1.3.1 Input

- $p_H, p_L, p_C$: true scenario parameters for three arms.
- $N_{H2}, N_{L2}, N_{C2}$: sample sizes of three arms in phase 2.
- $N_T^3, N_C^3$: sample sizes of treatment and control arms phase 3.
- $\delta_0, \eta_1, \eta_2, \xi$: parameters in Go/No-Go and Selection decisions.
- $z.test.\alpha$: parameter for the final decision, a nominal significance level (or say the corresponding critical value) for the final hypothesis test in phase 3.
- $\alpha_{H0}, \beta_{H0}, \alpha_{L0}, \beta_{L0}, \alpha_{C0}, \beta_{C0}$: parameters of prior distributions of the response rate
- Number of simulated trials

9.1.3.2 Output

- Decision table
- Probability of Go decision, probability of high-dose selection and Power
9.1.3.3 An Example (Figure 9.1)

Figure 9.1: An Example: Phase II/III Seamless Design with Binary Outcome
10. Phase II/III Seamless Designs for Continuous Outcome

Laiya extended the simulation scheme introduced in Thall and Simon (1994) for phase 2-3 seamless design, including three arms in the phase 2 stage, two doses (high and low) and placebo. At the end of the phase 2 stage, the design will make a go/no-go decision based on Bayesian go/no-go criteria first and then select one of high and low doses as the treatment arm in phase 3 based on Bayesian selection criteria.

10.1 Continuous Outcome

With continuous outcome, a two-sided hypothesis \( t.test \) will be performed based on data of the selected treatment dose integrated from both phase 2 and 3 and data of control arm only from phase 3 at the end of phase 3 stage,

\[
H_0 : \mu_T = \mu_C \quad \text{vs} \quad H_1 : \mu_T \neq \mu_C,
\]

where \( \mu_T \) and \( \mu_C \) represent response of the selected treatment dose and placebo respectively.

10.1.1 Model

In this section, these subscripts, \( H, L, C \) represent high dose, low dose and control respectively. For simplicity, here we suppress the subscript \( H, L \) and \( C \).

Let \( y_i \) be the response from the \( i \)th sampled subject of a certain arm in phase 2, \( i = 1, \cdots, N_2 \).

It is assumed that \( y_i \)'s are independent and identically distributed (i.i.d.) normal random variables, \( y_i \sim N(\mu, \sigma^2) \). And \( Y = (y_1, \cdots, y_{N_2}) \).

The likelihood is,

\[
y_i \sim N(\mu, \frac{1}{\tau});
\]
Priors are,

\[ \mu | \tau \sim N(\mu_0, \frac{1}{c_0 \tau}); \]
\[ \tau \sim Gamma(\alpha_0, \beta_0); \]

The conditional posterior distribution of \( \mu \) is

\[ \mu | Y, \tau \sim N\left( \frac{n \tau}{n \tau + c_0 \tau} \bar{y} + \frac{c_0 \tau}{n \tau + c_0 \tau} \mu_0, \frac{1}{n \tau + c_0 \tau} \right), \]

where \( \bar{y} = \sum_{i=1}^{N_2} y_i / N_2 \).

The conditional posterior distribution of \( \mu \) and \( \tau \) for priors is,

\[ P(\tau, \mu | Y) \propto P(\tau) p(\mu | \tau) P(Y | \mu, \tau) \]
\[ \propto \tau^{\alpha_0 - 1} e^{-\beta_0 \tau} \tau^{1/2} e^{\exp\left(-\frac{c_0 \tau}{2}(\mu - \mu_0)^2\right)} \tau^{n/2} e^{\exp\left(-\frac{\tau}{2} \sum (y_i - \bar{y})^2\right)} . \]

After integrating out \( \mu \), we get a Gamma posterior for \( \tau \),

\[ \tau | Y \sim Gamma(\alpha_0 + N_2/2, \beta_0 + \frac{1}{2} \sum (y_i - \bar{y})^2 + \frac{N_2 c_0}{2(N_2 + c_0)} (\bar{y} - \mu_0)^2). \]

By the above progress, we get \( \mu_H | Y_H, \tau_H, \mu_L | Y_L, \tau_L \) and \( \mu_C | Y_C, \tau_C \). It is assumed that three arms are independent, so

\[ \mu_H - \mu_C | Y_H, \tau_H, Y_C, \tau_C, \mu_L - \mu_C | Y_L, \tau_L, Y_C, \tau_C \quad \text{and} \quad \mu_H - \mu_L | Y_H, \tau_H, Y_L, \tau_L \]

are following normal distributions.

### 10.1.2 Decision Criteria

Bayesian go/no-go and selection decision with binary outcome from phase 2 that Laiya designed are as below:

- **Bayesian go/no-go criteria** at the end of phase 2 stage based on two indicators, \( h.go \) and \( l.go \). If \( h.go = 1 \) or \( l.go = 1 \), go to phase 3. Otherwise, not go.

Let \( h = Pr(\mu_H > \mu_C + \delta_0 | \text{data}) \) and \( l = Pr(\mu_L > \mu_C + \delta_0 | \text{data}) \)

\[ h.go = \begin{cases} 
1, & \text{if } h \geq \eta_1 \\
\sim Bin(1, h.go.p), & \text{if } \eta_2 < h < \eta_1 \\
0, & \text{if } h \leq \eta_2 
\end{cases} \]
Module 10. Phase II/III Seamless Designs for Continuous Outcome

\[ l.go = \begin{cases} 
1, & \text{if } l \geq \eta_1 \\
\sim \text{Bin}(1, l.go.p), & \text{if } \eta_2 < l < \eta_1 \\
0, & \text{if } l \leq \eta_2 
\end{cases} \]

where \( h.go.p = \frac{h-\eta_2}{\eta_1-\eta_2} \) and \( l.go.p = \frac{l-\eta_2}{\eta_1-\eta_2} \) and \( \delta_0 \) denotes the expected difference between the responses of treatment dose and placebo.

- **Bayesian selection criteria** after making go decision based on one indicator, \( h.select \). If \( h.select = 1 \), select the high dose \( (T = H) \). Otherwise, select the low dose \( (T = L) \).

\[ h.select = \begin{cases} 
1, & \text{if } Pr(\mu_H > \mu_L|data) > \xi \\
0, & \text{if } Pr(\mu_H > \mu_L|data) \leq \xi 
\end{cases} \]

- (Criteria of \( t.test \)) Let \( x_i \) be the response from the \( i \)th sampled subject of a certain arm in phase 3, \( i = 1, \cdots, N_3 \). It is assumed that \( x_i \)'s are independent and identically distributed (i.i.d.) normal random variables, \( x_i \sim N(\mu, \sigma^2) \). And \( X = (x_1, \cdots, x_{N_3}) \). Therefore, \( (Y_T, X_T) \) denotes responses of the selected treatment dose both in phase 2 and 3 and \( X_C \) responses of the control arm in phase 3. A two-sided \( t.test \) with significance level of \( t.test.\alpha \) will be performed based on \( (Y_T, X_T) \) and \( X_C \).

### 10.1.3 Program input and output

#### 10.1.3.1 Input

- \( \mu_H, \sigma_H, \mu_L, \sigma_L, \mu_C, \sigma_C \): true scenario parameters for three arms (see 2. Model).
- \( N_{H2}, N_{L2}, N_{C2} \): sample sizes of three arms in phase 2.
- \( N_{T3}, N_{C3} \): sample sizes of treatment and control arms phase 3.
- \( \delta_0, \eta_1, \eta_2, \xi \): parameters in Go/No-Go and Selection decisions (see 1. Introduction).
- \( t.test.\alpha \): parameter for the final decision, a nominal significance level (or say the corresponding critical value) for the final hypothesis test in phase 3.
- Priors:
  - \( \mu_{H0}, c_{H0}, \mu_{L0}, c_{L0}, \mu_{C0}, c_{C0} \) for the treatment effect of three arms.
  - \( \alpha_0, \beta_0 \) for the variance of treatment effect (here we assume three arms have the same prior distribution of variance).
10.1. Continuous Outcome

10.1.3. Program input and output

- Number of simulated trials

10.1.3.2 Output

- Decision table
- Probability of Go decision, Probability of dose selection, Power

10.1.3.3 An Example (Figure 10.1)

**Figure 10.1: An Example: Phase II/III Seamless Design with Continuous Outcome**
Part V

Master Protocols
11. Basket Trial Designs

11.1 Introduction

Basket trials are a type of master protocol in which a treatment is evaluated in more than one indications (baskets). For example, a BRAF inhibitor can be tested simultaneously in multiple cancer types all harboring BRAF mutations (Hyman et al., 2015) in a single trial (NCT01524978), as opposed to multiple trials each of which focusing on a single cancer type. Empowered by breakthroughs in genomics, complex diseases like cancer are further subdivided by biomarkers in addition to the histology, paving the foundation for complex studies like basket trials. In essence, a basket trial is a multi-arm phase 2 or phase 3 study investigating a treatment for multiple diseases or sub-diseases, and basket trials are usually without randomized control. Here and hereinafter, we use the terminology “basket” or “arm” to represent a group of patients with the same disease type or subtype that are treated by the same drug or drug combination in a multi-arm intervention trial.

Usually, each arm in a basket trial is compared with a historical control. Patients enrolled in a basket trial are often composed of a heterogeneous group across multiple indications, such as different cancer types. Therefore, it is difficult to evaluate time-to-event endpoints (e.g., progression-free survival (PFS) or overall survival (OS)), and the primary endpoints in a basket trial is often response rates (e.g., objective response rate (ORR) or pathological complete response (pCR)), which are less sensitive to the effects of population heterogeneity.

In screening new treatments, there might be a scientific rationale to expect some degree of similarity in treatment effect across arms. There exists two common approaches as to whether or not borrow information in the design and analysis of trial trial data: pooled analysis and independent analysis. If the treatment effect is assumed homogeneous across different baskets, a pooled analysis may be preferred, in which the data across all the arms are combined. However, the homogeneity assumption often fails in practice. For example, in BRAF V600 study, while BRAF V600E-mutant melanoma and hairy cell leukemia are responsive to BRAF inhibition, BRAF-mutant colon cancer
is not (Flaherty et al., 2010; Tiacci et al., 2011; Prahallad et al., 2012). When the homogeneity assumption is not valid, a separate stand-alone analysis for each arm is a simple alternative. However, conducting an independent evaluation in each arm is time- and resource-consuming. Also, the trial sample size may be inflated under independent arms when compared to designs that borrow information. Recently, adaptive designs that borrow information via model-based inference have been proposed, such as works in (Thall et al., 2003; Berry et al., 2013; Neuenschwander et al., 2016; Simon et al., 2016; Cunanan et al., 2017; Liu et al., 2017; Chu and Yuan, 2018a,b; Hobbs and Landin, 2018; Psioda et al., 2019). Using the observed data, these methods borrow information by prior distributions that shrink the arm-specific estimates to a centered value.

In U-Design, we implement a module of **Basket Trial Designs** and use simulation-based power calculation to evaluate four Bayesian approaches, including the Bayesian hierarchical model (BBHM) proposed by Berry et al. (2013), the calibrated Bayesian hierarchical model (CBHM) by Chu and Yuan (2018a), the exchangeabilitynonexchangeability (EXNEX) method in Neuenschwander et al. (2016) and a novel multiple cohort expansion (MUCE) method in Lyu et al. (2020). Users may choose a desirable designs based on provided software in this module.
Module 11. Basket Trial Designs

11.2 User Interface and Tutorial

11.2.1 Overview

Entering the Basket Trial Designs page, users will see two main tabs: Simulation Setup and Simulation Results. These two tabs allow users to conduct simulations and visualize/download simulation results. In the Simulation Setup tab, there are three steps (Figure 11.1): 1) Set trial parameters, 2) Select designs, and 3) Generate scenarios. Users need to complete the steps 1-3 to set up simulations for a single design or multiple designs. Upon completing steps 1-3, users click the “Launch Simulation” button at the bottom of the page. Users may also click the “Reset” button next to Launch Simulation to clear all the settings. After the simulations are launched, the results of simulations will be displayed in the Simulation Results tab. The simulation process can be monitored in real time at the top of the Simulation Results tab. Detailed steps of using this module are elaborated next in §11.2.2-§11.2.3.

Figure 11.1: Simulation Setup in the Basket Trial Designs module.
11.2. User Interface and Tutorial
11.2.2. Simulation Setup

In the Basket Trial Designs module, U-Design provides four designs, BBHM, CBHM, EXNEX, and MUCE, for simulation. Users can choose up to four design configurations for simultaneous comparison in the Simulation Setup tab each time. A design configuration means a design such as MUCE, along with the designs settings, such as sample size. Request to allow more than four design configurations by emailing admin@laiyaconsulting.com.

11.2.2.1 Step 1: Set trial parameters

Specify the number of simulated trials \( (n_{\text{sim}}) \) and the random seed of simulation \( (R_{\text{seed}}) \). Then select a number of arms \( (n_{\text{arm}}, 2 \leq n_{\text{arm}} \leq 10) \) from the dropdown box. Upon selection, manually type in the reference response rate \( (R_{\text{ref}}) \), the target response rate \( (R_{\text{target}}) \), and the type I error rate \( (\alpha) \) for each arm. See Figure 11.2.

Click the “Reset” button to clear all the settings. Users may click the \( \square \) icon (right after the cell of Arm 1) to copy and paste the value of Arm 1 into other arms, and click the \( \square \) icon (at the end of each row) to clear all the settings of the corresponding row.

Hover mouse over each trial parameter, and a description will be displayed explaining the meaning of the parameter. The detailed description of the above six input arguments is provided in Table 11.1.

Click the “Apply” button in Figure 11.2 to confirm and submit the trial parameters. And click the “Edit” button to enable the edits.

![Figure 11.2: Set trial parameters in the Basket Trial Designs module.](image)
Table 11.1: Input parameters for trials in the Basket Trial Designs module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{sim}$</td>
<td>Number of simulated trials</td>
<td>The number of simulated trials to be conducted for each scenario. The maximum number allowed is 10,000. Default value is 1,000.</td>
</tr>
<tr>
<td>$R_{seed}$</td>
<td>Random seed of simulation</td>
<td>A number used to initialize a pseudorandom number generator in the simulation. Default value is 32432.</td>
</tr>
<tr>
<td>$n_{arm}$</td>
<td>Number of arms</td>
<td>The number of arms in the trial. The range is $[2, 10]$.</td>
</tr>
<tr>
<td>$R_{ref}$</td>
<td>Reference response rate</td>
<td>The reference response rate (also called the historical control rate) is the largest rate considered to be not promising. Default value is 0.1.</td>
</tr>
<tr>
<td>$R_{target}$</td>
<td>Target response rate ($R_{target} &gt; R_{ref}$)</td>
<td>The target response rate is the smallest rate considered to be promising. Default value is 0.3.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Type I error rate</td>
<td>The probability of rejecting null when the null hypothesis is true. Default value is 0.1.</td>
</tr>
</tbody>
</table>
11.2.2.2 Step 2: Select designs

To select a design, click the button with the design’s name on it. Up to four design configurations may be selected for comparison. Upon selection of a design, specify the maximum sample size for each arm \((n)\), interim analysis parameters, and when needed, advanced design parameters. See Figure 11.3.

![Step 2: Select designs](image)

**Figure 11.3: Select designs** in the Basket Trial Designs module.

**Specify arm sample size**

U-Design provides a function to facilitate sample size specification. It generates “reference sample size” as candidate for simulations. Users can first try the reference sample size, generate simulation results, calibrate the sample size based on the results, and finally decide an appropriate sample size. Click the “Need help?” button in Figure 11.3 to expand the reference sample size section (Figure 11.4). U-Design provides three sets of sample sizes under power \((1 - \beta)\) of 70%, 80% and 90%, respectively, which are calculated by the one-sided equality Z-test with the standard
Module 11. Basket Trial Designs

development based on the target rate for one-sample proportion, \( n = \frac{(Z_{\alpha} + Z_{\beta})^2 R_{\text{target}}(1 - R_{\text{target}})}{(R_{\text{target}} - R_{\text{ref}})^2} \). Users can also manually type in a different power value and click the “Go” button to obtain a new reference sample size. These numbers can be used to help users to provide the maximum sample size for each arm. By clicking the \( \text{\textbullet} \) icon (at the end of each row), the sample sizes in the corresponding row will be loaded as the required maximum sample size. Click the “Hide the reference sample size” button to hide the reference sample size section. Similar in Step 1, users may click the \( \text{\textbullet} \) icon right after the cell of Arm 1 to copy and paste the sample size of Arm 1 into other arms, and may click the \( \text{\textbullet} \) icon at the end of the row to clear all sample size settings.

**Figure 11.4:** Display the reference sample size in Step 2: Select designs in the Basket Trial Designs module.

**Interim analysis (optional)**

Check the box behind the Optional: Include interim analysis in Figure 11.3 to expand the section of interim analysis parameters specification. Using the enrollment speed \( (S_{\text{enroll}}) \) of Arm 1 as a benchmark, users can manually type in the enrollment speeds for other arms that are relative to Arm 1. A value greater or less than 1 means a faster or slower patients accrual than Arm 1, respectively. Check the box of Apply futility stopping boundary and specify the probability threshold of futility stopping \( (P_{\text{futility}}) \) to allow interim analysis for futility.

When checked, two interim analyses will be applied by default. There are two possibilities. First, if all the arms are assumed to take the same amount of time to enroll the total number of patients (arm sample size) and the speed of enrollment is constant, the first interim analysis is
performed when each arm enrolls half (50%) of the sample size of the arm, and the second time is when each arm enrolls 75% of the total sample size. Otherwise, the first interim is conducted when the fastest arm enrolls half of the sample size of the arm, and the second interim is conducted when the slowest arm enrolls half of the sample size of the arm. For example, for a three-arms basket trial with the maximum sample size set at (40, 80, 20) for three arms, if the enrollment speed is $S_{enroll} = (1, 2, 0.5)$, the enrollment time of all three arms are the same. Assuming a constant enrollment speed, the two interim analyses will be performed when three arms enroll $(40 \times 0.5, 80 \times 0.5, 20 \times 0.5) = (20, 40, 10)$ patients and $(40 \times 0.75, 80 \times 0.75, 20 \times 0.75) = (30, 60, 15)$ patients, respectively; if the enrollment speed is $S_{enroll} = (1, 4, 0.75)$, two interim analyses will be performed when the fastest arm enrolls half patients (Arm 2) and the slowest arm enrolls half patients (Arm 1), which result in sample sizes $(10, 40, 3)$ for interim 1 and $(20, 80, 15)$ for interim 2. Request to allow other interim analysis options by emailing admin@laiyaconsulting.com.

**Design parameters**

The default values of advanced design parameters are recommended. See detailed explanation of each parameter in §11.3 next.

Click the “Apply” button in Figure 11.3 to confirm and submit the trial parameters. Click the “Edit” button to enable the edit mode and all design parameters can be modified. Click the “Delete” button to remove the selected designs.

Hover mouse over each design parameter, and a description will be displayed explaining the meaning of the parameter. The detailed description of the above input arguments is provided in Table 11.2 below.

**Table 11.2:** Input parameters for designs in the Basket Trial Designs module.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Maximum sample size</td>
<td>The maximum number of patients to be treated in the trial for each arm. The value is an integer between (0, 1000].</td>
</tr>
<tr>
<td>$S_{enroll}$</td>
<td>Relative enrollment speed</td>
<td>The enrollment speed relative to Arm 1. The range is $(0, \infty)$. Default value is 1 for all arms, which means all arms have the same enrollment speed. A value of 0.5 means the arm enrolls half of the speed of Arm 1, whatever it is.</td>
</tr>
<tr>
<td>$P_{futility}$</td>
<td>Futility stopping threshold</td>
<td>The probability threshold of futility stopping at an interim analysis. See stopping criteria in §11.3. Default value is 0.1.</td>
</tr>
</tbody>
</table>
11.2.2.3 Step 3: Generate scenarios

There are two ways to generate scenarios, automatically (in below Auto Generation tab, see Figure 11.5) or through manual construction (in below Manual Construction tab, see Figure 11.6).

Auto Generation (Figure 11.5)
Click the “Generate” button to automatically create three to six scenarios, each of which contains the true response rates for $n_{arm}$ arms. Scenario 1 is a global null scenario in which all arms are not promising with the response rate set at the reference response rate $R_{\text{ref}}$. Scenario 2 is a global alternative scenario in which all arms are promising with the response rate set at the target response rate $R_{\text{target}}$. Other scenario(s) are mixed scenarios with some but not all arms promising.

![Automatically generate scenarios in the Basket Trial Designs module.](image)

Manual Construction (Figure 11.6)
Follow the instructions below to manually construct scenarios. Then click the “Add” button to create these scenarios. The format of input must comply with the following instructions.

- Scenarios should be separated by linebreaks;
- Each scenario should contain $n_{arm}$ values, each representing the true response rate of each arm.
11.2. User Interface and Tutorial

11.2.2. Simulation Setup

- Response rate of each arm must be separated by a white space or comma.

For example, by inputting “0.1 0.2 0.2 0.1” or “0.1,0.2,0.2,0.1”, a scenario is presented with true response rates of four arms, 0.1, 0.2, 0.2 and 0.1. New created scenarios will be appended to the existing scenario list. If no scenarios are provided in the scenario list, the first newly-added scenario will be taken as the Null scenario, and the rest will be the alternative scenarios.

![Simulation Setup Interface](image.png)

**Figure 11.6**: Manually generate scenarios in the **Basket Trial Designs** module.

The generated scenarios are displayed as a list and editable (Figures 11.5 and 11.6) which appears below the generation section.

Click the “Delete” button (at the end of each row) to delete corresponding scenario. The first (Null) scenario is always included in order to benchmark designs. Click the “Delete All” button to delete all scenarios (including the Null scenario).
Module 11. Basket Trial Designs

11.2.2.4 Launch Simulation

Once the steps 1-3 are completed, users can conduct simulated clinical trials to examine the operating characteristics of the selected designs using the selected scenarios. Click the “Launch Simulation” button at the bottom of Simulation Setup tab (Figures 11.5 and 11.6). A “Launch Successful” message will be displayed on the screen (Figure 11.7) to indicate that the simulations have been successfully launched. Users may click the “Proceed To Simulation Results” button in the pop-up box to track the simulation processing status and simulation results.

Figure 11.7: “Launch Successful” message after launching simulation in the Basket Trial Designs module.
11.2.3 Simulation Results

In the Simulation Results tab, users can view the simulation progress and simulation results (§11.2.3.1), restore the simulation settings if needed (§11.2.3.2), and download U-Design’s proprietary intelligent simulation reports (§11.2.3.3). Specifically, all the simulation results (figures and tables) can be downloaded in Word format, accompanying the statistical sections in a trial protocol. Hereinafter, we use simulation results and operating characteristics interchangeably.

11.2.3.1 View simulation results

In the Simulation Results tab, the Running Simulations panel exhibits the progress of ongoing simulation (Figure 11.8). The ongoing simulations are displayed in ascending order by the launch time. Click the icon “×” to delete the corresponding simulation.

![Figure 11.8: Simulation progress in the Basket Trial Designs module.](image)

Once the simulations are completed, the Running Simulations panel in Figure 11.8 will disappear, green “simulation result created” massages will appear instead and stay at the same place of the Running Simulations panel unless explicitly dismissed by clicking the icon “×” at the end of the corresponding row, and the simulation results will be automatically loaded into the Simulation History panel (Figure 11.9), with the blue mail icon 💌 shown to indicate new results. All the previously completed simulations are also listed in the Simulation History panel. Simulation results for other modules can also be viewed under the Simulation History by dropping down the “Select a module” button (Figure 11.9). Click the 🗑️ button to delete the selected simulation results.
Module 11. Basket Trial Designs

Figure 11.9: Simulation Results in the Basket Trial Designs module.

Click the button to unfold the simulation results (Figure 11.10). The design settings are firstly displayed at the top of each simulation study. Then the results of simulation are shown in two ways: figures and tables. See next.

Figure 11.10: View the simulation results in the Basket Trial Designs module.


Simulation Results Summary (Figures 11.11 and 11.12)

There are two sections in the Simulation Results Summary.

1. Line plots showing three frequentist summary statistics of the simulation results for all the designs from two aspects: **Family-wise Type I Error Rate** and **Family-wise Power** (Figure 11.11).
   - The three frequentist summary statistics are explained in full detail next.
     - **Family-wise Type I Error Rate**: The proportion of simulated trials in which any true null is rejected, i.e., any false discovery is made. In other words, it is the proportion of simulated trials in which any arm is wrongly declared to be more efficacious than historical controls.
     - **Family-wise Power**: Two subtypes of powers are considered.
       * **Family-wise Power 1**: The proportion of simulated trials in which only true efficacious arms are correctly declared to be more efficacious than the historical controls, and no true inefficacious arms are wrongly declared to be more efficacious than the historical controls.
       * **Family-wise Power 2**: The proportion of simulated trials in which all true efficacious arms are correctly declared to be more efficacious than the historical controls, and no true inefficacious arms are declared to be more efficacious than the historical controls.
   - For each line plot, the x-axis is the index of scenario and the y-axis is the value of summary statistics. Lines with different colors represent different designs.
   - The plots are interactive for better visualization.
     - Hover the mouse on a dot and a box will display the value of each design at the corresponding scenario (e.g. top left plot in Figure 11.11: Family-wise Type I Error Rate).
     - Hover the mouse on the design label to highlight the corresponding line and fade the others (e.g. top right plot in Figure 11.11: Family-wise Power 1).
     - Click the design label to hide the corresponding line and click again to change it back (e.g. bottom left plot in Figure 11.11: Family-wise Power 2).

2. A table showing trial settings and probability thresholds used in the final analysis for all designs (Figure 11.12).
   - The table shows the trial parameters specified in step 1 (§11.2.2) and the probability of your trial designs anywhere, anytime.
thresholds for the rejection of null in the final analysis for all the selected designs. The trial parameters displayed include the reference response rate ($R_{\text{ref}}$), the target response rate ($R_{\text{target}}$), and the type I error rate ($\alpha$), for each arm.

Figure 11.12: Trial settings and probability thresholds for the final analysis in the Basket Trial Designs module.
Simulation Results by Scenario (Figure 11.13)

Full simulation results are presented in bar plots and tables arranged by scenario (Figure 11.13). For each scenario, the simulation results are summarized from the following three frequentist aspects.

1. **Type I error rate / Power:** A bar plot showing the arm-wise type I error rate & power and family-wise type I error rate & power (FWER & FW-power).
   - Bars with different colors represent different designs.
   - The first \( n_{arm} \) clusters of bars report the arm-wise type I error rate & power, and the last three clusters report the FWER and two family-wise powers.
   - Four statistics are explained in detail next.
     - **Arm-wise type I error rate & power:** The proportion of simulated trials in which the null hypothesis for an arm is rejected, i.e., the proportion of simulated trials in which the arm is declared to be more efficacious than the historical control. This is the arm-wise type I error rate if the arm is actually not more efficacious than the historical control in this arm, and is the arm-wise power otherwise.
     - **Family-wise type I error rate & power (FWER & FW-power)**
       - **Family-wise type I error rate (FWER):** The proportion of simulated trials in which at least one arm is wrongly declared to be more efficacious than historical controls in any arm.
       - **Family-wise power 1 (FW-power1):** The proportion of simulated trials in which only true efficacious arms are correctly declared to be more efficacious than the historical controls, and no true inefficacious arms are wrongly declared to be more efficacious than the historical controls.
       - **Family-wise power 2 (FW-power2):** The proportion of simulated trials in which all true efficacious arms are correctly declared to be more efficacious than the historical controls, and no true inefficacious arms are declared to be more efficacious than the historical controls.
   
   For detailed descriptions, please refer to Simulation Results Summary above.

2. **Response Rate Estimation:** A table is provided (Figure 11.13) reporting the accuracy and the precision of the estimates of response rates. The first two columns summarize the scenario settings, with the index and its true response rate of each arm; the subsequent columns report the average bias of response rate estimates and their standard deviation. The bias is defined as the difference between the posterior mean of response rate and the true response rate. The average is taken across all the simulated trials.
3. **Interim Analysis**: A table is provided (Figure 11.13) summarizing the statistics of interim analysis, if any.

- **Average sample size (s.d.)**: The average number of patients treated in a simulated trial and its standard deviation, averaging across all the simulated trials.
- **Current # of patients treated**: The numbers of patients treated for each arm when the 1st and the 2nd interim analyses are performed, respectively.
- **Probability of futility stopping**: The proportion of simulated trials in which an arm is stopped early due to futility at the 1st or the 2nd interim analysis.

When calculating the standard deviation, we use $n_{sim}$ as the denominator instead of $(n_{sim} - 1)$ in U-Design.

**Figure 11.13**: Simulation results by scenario in the Basket Trial Designs module.
11.2.3.2 Restore simulation setup

Users can restore the simulation settings from the simulation results by clicking the button at the upper right corner of each simulation results panel (yellow arrow in Figure 11.14) and the display will switch to the Simulation Setup page with the same simulation settings restored. This is useful to restore the old simulation settings for reproducible results.

![Simulation Setup](Image)

Figure 11.14: Restore simulation setup and download simulation results in the Basket Trial Designs module.

11.2.3.3 Download simulation results

A button is placed at the upper right corner of each simulation results panel (green arrow in Figure 11.14). Click it to download U-Design’s proprietary word file with complete simulation results under the designs and scenarios users specified in the simulation settings tab. Users could update the simulation settings and results tailored for their trials. Contact us via email (admin@laiyaconsulting.com) for consulting services.
11.3 Statistical Methods Review

11.3.1 Bayesian Hierarchical Model (BBHM)

Thall et al. (2003) apply a Bayesian hierarchical model to phase II basket trial designs that borrows information across arms.

11.3.1.1 Probability Model

Consider a phase II basket trial that evaluates the efficacy of a new treatment in $K$ different arms (indications). Let $n_k$ and $y_k$ denote the number of patients and responders in arm $k$, respectively. Denote by $p_k$ the true and unknown response rate for arm $k$. The objective of the trial is to test the null hypothesis that the response rate, $p_k$, of the arm is less than a reference response rate, $\pi_{k0}$,

$$H_{0k}: p_k \leq \pi_{k0}$$

versus the alternative hypothesis that the response rate is at least as high as a target rate, $\pi_{k1}$,

$$H_{1k}: p_k \geq \pi_{k1},$$

for each arm $k$, $k = 1, 2, \ldots, K$.

BBHM models the log-odds of response rate for each arm $k$, including an adjustment for the targeted $\pi_{k1}$ rates, defined as

$$\theta_k = \log\left(\frac{p_k}{1 - p_k}\right) - \log\left(\frac{\pi_{k1}}{1 - \pi_{k1}}\right).$$

Assume $\theta_k$ follow a normal prior distribution with unknown mean $\theta$ and variance $\sigma^2$

$$\theta_k \mid \theta \overset{iid}{\sim} N(\theta, \sigma^2).$$

The hyperparameters $\theta$ and $\sigma^2$ are given conjugate hyperpriors,

$$\theta \sim N(\theta_0, \sigma_0^2), \quad \sigma^2 \sim \text{Inv-Gamma}(\alpha_s, \lambda_s),$$

where $\alpha_s$ and $\lambda_s$ are the shape and scale parameters of the inverse gamma distribution, respectively. This prior construction assumes that the arm-specific treatment effect $\theta_k$’s across different arms are exchangeable and shrinks to a shared mean $\theta$, thus enabling information borrowing across arms. The degree of shrinkage or information borrowing is determined by the value of $\sigma^2$. The smaller the $\sigma^2$, the stronger the borrowing. In the extreme cases, $\sigma^2 = 0$ means all $\theta_k$’s equal $\theta$ which is
the pooled analysis, and \( \sigma^2 = \infty \) is equivalent to the independent approach, where \( \theta_k \) are assumed independent and distinct.

In short, the hierarchical models are:

\[
\begin{align*}
\text{Likelihood:} & \quad y_k | n_k, p_k \sim \text{Binomial}(n_k, p_k) \\
\text{Transformation:} & \quad \theta_k = \log \left( \frac{p_k}{1-p_k} \right) - \log \left( \frac{\pi_{k1}}{1-\pi_{k1}} \right) \\
\text{Prior for } \theta_k : & \quad \theta_k | \theta, \sigma^2 \sim N(\theta, \sigma^2) \\
\text{Hyperpriors:} & \quad \theta \sim N(\theta_0, \sigma^2_0) \\
& \quad \sigma^2 \sim \text{Inv-Gamma}(\alpha_s, \lambda_s)
\end{align*}
\]

Following Berry et al. (2013), by default, U-Design assigns a non-informative inverse gamma prior Inv-Gamma(0.0005, 0.000005) for \( \sigma^2 \), and uses the average of \( \theta_k \) under the null rates \( \theta_0 = \frac{1}{K} \sum_{k=1}^{K} \left( \log \left( \frac{\pi_{k0}}{1-\pi_{k0}} \right) - \log \left( \frac{\pi_{k1}}{1-\pi_{k1}} \right) \right) \) and a large variance \( \sigma^2_0 = 10^2 \) for the prior of \( \theta \), creating a nearly non-informative prior. The inverse gamma prior gives a \( E(\sigma^2) = 10^2 \) and \( \text{Var}(\sigma^2) = 2 \times 10^7 \).

### 11.3.1.2 Trial Design

Suppose \( L(\geq 0) \) interim looks are planned, and the \( l \)-th interim analysis is conducted after \( n^l_k \) patients have been enrolled in arm \( k \). Let \( D^l \equiv \{(n^l_k, y^l_k) : k = 1, 2, \ldots, K\} \) denote the observed data at interim analysis \( l \), where \( y^l_k \) is the number of responders among the \( n^l_k \) patients. Denote \( D^{L+1} \equiv \{(n^{L+1}_k, y^{L+1}_k) : k = 1, 2, \ldots, K\} \) the observed data at the end of the trial, where \( n^{L+1}_k \) is the prespecified maximum sample size for arm \( k \) and \( y^{L+1}_k \) is the total number of responders. The proposed BBHM basket trial design with \( L \) interim looks is describe as follows:

1. Enroll \( n^1_k \) patients in \( k \)-th arm, \( k = 1, 2, \ldots, K \).
2. Given the data \( D^l \) at the \( l \)-th interim look, \( l = 1, 2, \ldots, L \),
   
   (a) [Futility stopping] If the posterior probability that the response rate of arm \( k \), \( p_k \), is greater than \( (\pi_{k0} + \pi_{k1})/2 \) is small, i.e.,
   \[
   Pr\{p_k > \frac{\pi_{k0} + \pi_{k1}}{2} \mid D^l\} < P_{\text{futility}},
   \]
   stop the accrual to the \( k \)-th arm for futility;
   
   (b) Otherwise, continue to enroll patients until reaching the next interim analysis.

Your trial designs anywhere, anytime.
3. Once the maximum sample size is reached or all the arms have stopped, evaluate the efficacy for each arm based on all the observed data. If the posterior probability that the response rate, \( p_k \), is greater than \( \pi_{k0} \) is large, i.e.,

\[
Pr\{p_k > \pi_{k0} \mid D_{i+1}^L\} > \phi_k,
\]

arm \( k \) is declared efficacious and promising; otherwise, it is considered not promising.

Step 2 is optional, since the BBHM design does not require an interim look. However, it is useful to allow interim in practice for early stopping. The probability thresholds for the interim analysis \( P_{\text{futility}} \) and for the final analysis \( \{\phi_k : k = 1, 2, \ldots, K\} \), are calibrated through simulations to achieve a prespecified type I error rate for each arm under the global null scenario. In brief, assume \( n_{sim} \) trials are simulated under the Null scenario. For arm \( k \), suppose \( T_k \) out of \( n_{sim} \) trials are early stopped due to futility. From the remaining \( (n_{sim} - T_k) \) trials, we can obtain \( (n_{sim} - T_k) \) posterior probabilities \( p(p_k > \pi_{k0} \mid H_0) \). Denote them as \( \{P_i = Pr\{p_k > \pi_{k0} \mid D_{i+1}^L\}, i = 1, \ldots, n_{sim} - T_k\} \), where \( D_{i+1}^L \) is the observed data at the end of \( i \)-th trial under the null scenario. Then sort the samples \( \{P_i\} \) to obtain a set of order statistics \( \{P_{(i)} : i = 1, \ldots, n_{sim} - T_k\} \), where \( P_{(i)} \leq P_{(j)} \), for \( i < j \). Finally, \( \phi_k = P_{(n_{sim} - T_k - n_{sim} \times \alpha_k)} \) so that \( n_{sim} \times \alpha_k \) out of \( n_{sim} \) trials are rejected under the Null scenario, i.e., the type I error rate is \( \alpha_k \).
11.3. Statistical Methods Review

11.3.2. Calibrated Bayesian Hierarchical Model (CBHM)

Chu and Yuan (2018a) proposed a calibrated Bayesian hierarchical model (CBHM) as an extension of BBHM, which estimates \( \sigma^2 \) from the observed data instead of using a prior.

11.3.2.1 Probability Model

Consider a phase II basket trial that evaluates the efficacy of a new treatment in \( K \) different arms (indications). Let \( p_k \) denote the true and unknown response rate for arm \( k \). The objective of the trial is to test whether the new treatment is effective in each of the arms

\[
H_{0k} : p_k \leq \pi_{k0} \quad \text{versus} \quad H_{1k} : p_k \geq \pi_{k1}, \quad \text{for } k = 1, 2, \ldots, K,
\]

where \( \pi_{k0} \) is the reference response rate (also called the historical response rate), and \( \pi_{k1} \) is the target response rate under which the treatment is regarded as promising.

Suppose at a certain moment, \( n_k \) patients from arm \( k \) have been enrolled, among which \( y_k \) patients respond favorably to the treatment. CBHM assumes that \( y_k \) follows a hierarchical model

\[
\begin{align*}
\text{Likelihood:} & \quad y_k \mid n_k, p_k \sim \text{Binomial}(n_k, p_k) \\
\text{Transformation:} & \quad \theta_k = \log \left( \frac{p_k}{1 - p_k} \right) \\
\text{Prior for } \theta_k : & \quad \theta_k \mid \theta, \sigma^2 \sim N(\theta, \sigma^2) \\
\text{Hyperpriors:} & \quad \theta \sim N(\theta_0, \sigma_0^2)
\end{align*}
\]

The same as Berry et al. (2013), the above prior construction assumes that the arm-specific treatment effect \( \theta_k \)'s across different arms are exchangeable and shrinks to a shared mean \( \theta \), thereby enabling information borrowing across arms. The degree of shrinkage or information borrowing is determined by the value of \( \sigma^2 \). Following Chu and Yuan (2018a), by default, U-Design uses the average of \( \theta_k \) under the null rates \( \theta_0 = \frac{1}{K} \sum_{k=1}^{K} \log \left( \frac{\pi_{k0}}{1 - \pi_{k0}} \right) \) and a large variance \( \sigma_0^2 = 10^2 \) for the prior of \( \theta \), creating a vague prior.

11.3.2.2 Calibration of shrinkage parameter \( \sigma^2 \)

Unlike the BBHM approach (Berry et al., 2013) in §11.3.1, which assigns a prior to \( \sigma^2 \) and estimates it from the data, CBHM defines \( \sigma^2 \) in (11.1) as a function of the measure of homogeneity among the arms. The idea is that the function is prespecified and calibrated in a way such that when the treatment effects across arms are homogeneous, small \( \sigma^2 \) is induced so that strong information borrowing occurs and thus improves power, and when the treatment effects across arms...
are heterogeneous, large $\sigma^2$ is induced so that little or no borrowing across groups occur, thereby controlling the type I error rate. In what follows, Chu and Yuan (2018a) use a homogeneity measure to determine and calibrate the estimation of parameter $\sigma^2$.

Specifically, CBHM adopts the chi-squared test statistic to measure homogeneity, given by

$$T = \sum_{k=1}^{K} \frac{(O_{0k} - E_{0k})^2}{E_{0k}} + \sum_{k=1}^{K} \frac{(O_{1k} - E_{1k})^2}{E_{1k}}$$

where $O_{0k}$ and $O_{1k}$ denote the observed counts of non-responses and responses for arm $k$ (i.e. $n_k - y_k$ and $y_k$), and $E_{0k}$ and $E_{1k}$ are the “expected” counts of non-responses and responses, given by

$$E_{0k} = n_k \frac{\sum_k n_k - \sum_k y_k}{\sum_k n_k} \quad \text{and} \quad E_{1k} = n_k \frac{\sum_k y_k}{\sum_k n_k}$$

A smaller value of $T$ indicates higher homogeneity in the treatment effect across arms.

Then CBHM links the shrinkage parameter $\sigma^2$ with $T$ through the following two-parameter exponential model

$$\sigma^2 = g(T) = \exp\{a + b \times \log(T)\}, \quad (11.2)$$

where $a$ and $b$ are tuning parameters that characterize the relationship between $\sigma^2$ and $T$. Also $b > 0$ is required so that greater homogeneity (i.e. a small value of $T$) leads to stronger shrinkage (i.e. a small value of $\sigma^2$). The values of $a$ and $b$ in (11.2) are calibrated using the following three-step simulation-based procedure:

1. Simulate the case in which the treatment is effective for all arms. Specifically, $R$ replicates of data are generated by simulating $y = (y_1, \ldots, y_K)$ from $\text{Binomial}(n, \pi_1)$, where $n = (n_1, \ldots, n_K)$ and $\pi_1 = (\pi_{11}, \ldots, \pi_{K1})$ and then calculate $T$ for each simulated dataset. Let $H_{B1}$ denote the median of $T$ from $R$ simulated datasets.

2. Simulate the cases in which the treatment effect is heterogeneous across arms. Let $\pi(k) = (\pi_{11}, \ldots, \pi_{k1}, \pi_{(k+1)0}, \ldots, \pi_{K0})$ denote scenario in which the treatment is effective for the first $k$ arms with the target response rate of $\pi_{k1}$, but not effective for arms $(k + 1)$ to $K$ with the reference response rate of $\pi_{k0}$. Given a value of $k$, we generate $R$ replicates of data by simulating $y$ from $\text{Binomial}(n, \pi(k))$, calculate $T$ for each simulated dataset and then obtain its median $H_{B2k}$. Repeat this for $k = 1, 2, \ldots, K - 1$ and define

$$H_{B2} = \min_k (H_{B2k})$$

3. Let $\sigma^2_{B1}$ denote a prespecified small value (the default value is 1 in U-Design) for shrinkage parameter $\sigma^2$ under which strong shrinkage or information borrowing occurs under the hierarchical model (equation (11.1)), and let $\sigma^2_{B2}$ denote a prespecified large value (the default value
is 80 in U-Design) of shrinkage parameter $\sigma^2$, under which little shrinkage or information borrowing occurs. Solve $a$ and $b$ in equation (11.2) based on the following two equations

\[
\begin{align*}
\sigma^2_{B1} &= g(H_{B1}; a, b) = \exp\{a + b \times \log(H_{B1})\} \\
\sigma^2_{B2} &= g(H_{B2}; a, b) = \exp\{a + b \times \log(H_{B2})\}
\end{align*}
\]  

(11.3)

which enforces strong and weak shrinkage respectively. The solution of the equations (11.3) is given by

\[
\begin{align*}
a &= \log(\sigma^2_{B1}) - \frac{\log(\sigma^2_{B2}) - \log(\sigma^2_{B1})}{\log(H_{B2}) - \log(H_{B1})} \log(H_{B1}) \\
b &= \frac{\log(\sigma^2_{B2}) - \log(\sigma^2_{B1})}{\log(H_{B2}) - \log(H_{B1})}
\end{align*}
\]

U-Design’s take: While we report the procedure from Chu and Yuan (2018a), we leave the users to assess the procedure in §11.3.2.2. We would probably take a formal empirical Bayes approach instead, such as the procedure in Carlin and Louis (2010).

**11.3.2.3 Trial Design**

CBHM applies the same trial design as that in BBHM (§11.3.1).
11.3.3 ExchangeabilityNonexchangeability (EXNEX) Method

Neuenschwander et al. (2016) proposed the exchangeabilitynonexchangeability (EXNEX) approach that allows each arm-specific parameter to be exchangeable with other similar arm parameters or nonexchangeable with any of them.

11.3.3.1 Probability Model

Consider a phase II basket trial that evaluates the efficacy of a new treatment in \( K \) different arms (indications). Let \( n_k \) and \( y_k \) denote the number of patients and responders in arm \( k \), respectively. Denote by \( p_k \) the true and unknown response rate for arm \( k \). A natural sampling model for \( y_k \) given \( n_k \) and \( p_k \) is binomial model, \( y_k | n_k, p_k \sim \text{Binomial}(n_k, p_k) \).

The objective of the trial is to test whether the new treatment is effective in each of the arms

\[
H_{0k} : p_k \leq \pi_{k0} \quad \text{versus} \quad H_{1k} : p_k \geq \pi_{k1},
\]

for \( k = 1, 2, \ldots, K \), where \( \pi_{k0} \) and \( \pi_{k1} \) are the reference and target response rates for arm \( k \), respectively. Let \( \theta_k = \log \left( \frac{p_k}{1-p_k} \right) \) denote the log-odds of the response rate. EXNEX models the \( \theta_k \)'s with a mixture distribution,

\[
\theta_k \mid w_k, \theta_{\text{EX},c}, \sigma_{\text{EX},c}^2, \sigma_{\text{NEX},k}^2, \sigma_{\text{NEX},k}^2 \sim \sum_{c=1}^{C} w_{kc} N(\theta_{\text{EX},c}, \sigma_{\text{EX},c}^2) + w_{k0} N(\theta_{\text{NEX},k}, \sigma_{\text{NEX},k}^2), \quad (11.4)
\]

In other words, with probability \( w_{kc} \), \( \theta_k \) belongs to an exchangeability (EX) component \( c \), and with probability \( w_{k0} \), \( \theta_k \) belongs to a nonexchangeability (NEX) component. Here, \( \sum_{c=0}^{C} w_{kc} = 1 \).

The parameters of the EX components, \( \theta_{\text{EX},c} \) and \( \sigma_{\text{EX},c}^2 \) are shared across arms within component \( c \). In contrast, the parameter of the NEX components, \( \theta_{\text{NEX},k} \) and \( \sigma_{\text{NEX},k}^2 \) are arm-specific. The number of EX components \( C \) and the weights of the components \( w_k = (w_{k1}, \ldots, w_{kC}, w_{k0}) \) are prespecified by the investigator. By default, the same NEX components and mixture weights are specified for all arms, \( \theta_{\text{NEX},1} = \ldots = \theta_{\text{NEX},K} = \theta_{\text{NEX}}, \sigma_{\text{NEX},1}^2 = \ldots = \sigma_{\text{NEX},K}^2 = \sigma_{\text{NEX}}^2 \), and \( w_1 = \ldots = w_K = w \). For the prior specification, in each EX component \( c \), a normal prior is assigned to \( \theta_{\text{EX},c} \), and a half-normal (HN) prior with scale parameter \( s_c \) is assigned to \( \sigma_{\text{EX},c} \).

\[
\theta_{\text{EX},c} \sim N(\mu_{\text{EX},c0}, \sigma_{\text{EX},c0}^2), \quad \sigma_{\text{EX},c} \sim \text{HN}(s_c).
\]

In U-Design, the default settings Neuenschwander et al. (2016) is used for EXNEX: A mixture of two \( (C = 2) \) EX distributions and one NEX distribution with weights \( w = (0.25, 0.25, 0.5) \) is...
11.3. Statistical Methods Review

11.3.3. Exchangeability Nonexchangeability (EXNEX) Method

chosen by default. Therefore, in brief, U-Design applies the following hierarchical model:

\[
\begin{align*}
\text{Likelihood:} & \quad y_k \mid n_k, p_k \sim \text{Binomial}(n_k, p_k) \\
\text{Transformation:} & \quad \theta_k = \log \left( \frac{p_k}{1 - p_k} \right) \\
\text{Prior for } \theta_k : & \quad \theta_k \mid w, \theta_{EX}, \sigma_{EX}^2, \theta_{NEX}, \sigma_{NEX}^2 \sim \\
& \quad 0.25 N(\theta_{EX,1}, \sigma_{EX,1}^2) + 0.25 N(\theta_{EX,2}, \sigma_{EX,2}^2) + 0.5 N(\theta_{NEX}, \sigma_{NEX}^2) \\
\text{Hyperpriors:} & \quad \theta_{EX,1} \sim N(\mu_{EX,10}, \sigma_{EX,10}^2), \sigma_{EX,1} \sim \text{HN}(s_1) \\
& \quad \theta_{EX,2} \sim N(\mu_{EX,20}, \sigma_{EX,20}^2), \sigma_{EX,2} \sim \text{HN}(s_2)
\end{align*}
\]

Following Neuenschwander et al. (2016), weakly-informative priors are used in U-Design by default. Specifically, for the priors of the NEX parameters, we fix the mean \( \theta_{NEX} \) at the log-odds of a plausible guess for the response probability (e.g. the mean of the middle of reference and target response rates across arms, \( p_w = \frac{1}{K} \sum_{k=1}^{K} \pi_k + \pi_{10} \)), and the variance \( \sigma_{NEX}^2 \) at a value that corresponds to approximately one observation, \( \sigma_{NEX}^2 = 1/p_w + 1/(1 - p_w) \), for all arms. For EX components, we place \( N \left( \log \left( \frac{\pi_0}{1 - \pi_0} \right), 1/\pi_0 + 1/(1 - \pi_0) - 1 \right) \) and \( N \left( \log \left( \frac{\pi_1}{1 - \pi_1} \right), 1/\pi_1 + 1/(1 - \pi_1) - 1 \right) \) prior on \( \theta_{EX,1} \) and \( \theta_{EX,2} \), respectively, where \( \pi_0 = \frac{1}{K} \sum_{k=1}^{K} \pi_{k0} \) and \( \pi_1 = \frac{1}{K} \sum_{k=1}^{K} \pi_{k1} \) are the average reference and target response rate across arms; and half-normal priors with scale parameter \( s_1 = s_2 = 1 \) on \( \sigma_{EX,1} \) and \( \sigma_{EX,2} \).

11.3.3.2 Trial Design

The original EXNEX design does not have a futility or efficacy stopping rule, but for fair comparison, the same rules as those in BBHM (§11.3.1) are available in U-Design.

*Your trial designs anywhere, anytime.*
11.3.4 Multiple Cohort Expansion (MUCE) Method

The multiple cohort expansion (MUCE) design (Lyu et al., 2020) is originally proposed by Laiya Consulting Inc, for trials with multiple arms, include basket trials. The MUCE is based on a class of Bayesian hierarchical models including a latent probit prior that allows for different degrees of borrowing across arms. Furthermore, instead of using the posterior interval of the estimated response rate to declare futility or efficacy, as in BBHM (§11.3.1), CBHM (§11.3.2) and EXNEX (§11.3.3), MUCE applies a formal Bayesian hypothesis test to make statistical inference.

11.3.4.1 Probability Model

Consider a phase II basket trial that evaluates the efficacy of a new treatment in \(K\) different arms (indications). Suppose \(n_k\) patients have been treated in arm \(k\), and \(y_k\) of them respond. Let \(p_k\) denote the true and unknown response rate for the arm \(k\). We assume \(y_k\) follows a binomial distribution conditional on \(n_k\) and \(p_k\), \(y_k | n_k, p_k \sim \text{Binomial}(n_k, p_k)\). Whether arm \(k\) is effective can be examined by the following hypothesis test:

\[
H_{0k} : p_k \leq \pi_{k0} \quad \text{versus} \quad H_{1k} : p_k > \pi_{k0},
\]

(11.6)

where \(\pi_{k0}\) is the reference response rate for arm \(k\).

MUCE constructs a formal Bayesian testing framework for (11.6). Let \(\lambda_k\) be a binary indicator of the hypothesis, such that \(\lambda_k = 0\) (or 1) represents that hypothesis \(H_{0k}\) (or \(H_{1k}\)) is true. Firstly, a prior model for \(p_k\) is built under each hypothesis. Let \(\theta_k = \log \left( \frac{p_k}{1-p_k} \right)\) denote the log-odds of the response rate. The null hypothesis \(p_k \leq \pi_{k0}\) is equivalent to \(\theta_k \leq \theta_{k0}\), and the alternative hypothesis is equivalent to \(\theta_k > \theta_{k0}\), where \(\theta_{k0} = \log \left( \frac{\pi_{k0}}{1-\pi_{k0}} \right)\). Conditional on \(\lambda_k\), MUCE assumes

\[
\theta_k \mid \lambda_k = 0 \sim \text{Trunc-Cauchy}(\theta_{k0}, \gamma; (-\infty, \theta_{k0}]),
\]

\[
\theta_k \mid \lambda_k = 1 \sim \text{Trunc-Cauchy}(\theta_{k0}, \gamma; (\theta_{k0}, \infty)),
\]

where Trunc-Cauchy(\(\theta, \gamma; A\)) denotes a Cauchy distribution with location \(\theta\) and scale \(\gamma\) truncated to interval \(A\).

Secondly, prior models for the probabilities of the hypotheses (i.e. priors for the probabilities of \(\{\lambda_k = 1\}\)) are constructed. MUCE uses a probit model as the prior model for \(\lambda_k\). Let \(Z_k\) be a latent Gaussian random variable, and \(\lambda_k = I(Z_k < 0)\), where \(I(\cdot)\) is an indicator function. \(Z_k\) is assumed to follow a normal distribution,

\[
Z_k \sim N(\eta_k, \sigma_0^2).
\]
Here, $E(Z_k) = \eta_k$, in which $\eta_k$ characterizes the effect of arm $k$. The arm-specific effects are then separately modeled by common priors,

$$\eta_k \mid \eta_0, \sigma^2_\eta \sim N(\eta_0, \sigma^2_\eta).$$

Lastly, give $\eta_0$ a hyperprior, $\eta_0 \sim N(\mu_{\eta_0}, \sigma^2_{\eta_0})$.

In brief, the entire hierarchical models are summarized in the following display:

Likelihood: $y_k \mid n_k, p_k \sim \text{Binomial}(n_k, p_k)$;

Transformation: $\theta_k = \log \left( \frac{p_k}{1-p_k} \right)$, $\theta_k0 = \log \left( \frac{\pi_k0}{1-\pi_k0} \right)$;

Prior for $(\theta_k \mid \lambda_k)$: $\theta_k \mid \lambda_k = 0 \sim \text{Trunc-Cauchy}(\theta_k0, \gamma; (-\infty, \theta_k0])$,
$$\theta_k \mid \lambda_k = 1 \sim \text{Trunc-Cauchy}(\theta_k0, \gamma; (\theta_k0, \infty));$$

Prior for $\lambda_k$: $\lambda_k = \begin{cases} 0, & \text{if } Z_k < 0, \\ 1, & \text{if } Z_k \geq 0; \end{cases}$ (11.7)

Latent probit regression: $Z_k \mid \eta_k, \sigma^2_0 \sim N(\eta_k, \sigma^2_0)$;

Arm-specific effects: $\eta_k \mid \eta_0, \sigma^2_\eta \sim N(\eta_0, \sigma^2_\eta)$;

Hyperprior: $\eta_0 \mid \mu_{\eta_0}, \sigma^2_{\eta_0} \sim N(\mu_{\eta_0}, \sigma^2_{\eta_0})$.

In U-Design, the values of the hyperparameters $\gamma = 2.5$, $\mu_{\eta_0} = 0$, $\sigma^2_0 = 100$, $\sigma^2_\eta = 1$ and $\sigma^2_{\eta_0} = 1$ are used by default.

### Trial Design

Suppose $L (\geq 0)$ interim looks are planned, and the $l$-th interim analysis is conducted after $n_k^l$ patients have been enrolled in arm $k$. Let $D^l = \{(n_k^l, y_k^l) : k = 1, 2, \ldots, K\}$ denote the observed data at interim analysis $l$, where $y_k^l$ is the number of responders among the $n_k^l$ patients. Denote $D^{L+1} = \{(n_k^{L+1}, y_k^{L+1}) : k = 1, 2, \ldots, K\}$ the observed data at the end of the trial, where $n_k^{L+1}$ is the prespecified maximum sample size for arm $k$ and $y_k^{L+1}$ is the total number of responders. The proposed phase II basket trial design with $L$ interim looks is describe as follows:

1. Enroll $n_k^1$ patients in $k$-th arm, $k = 1, 2, \ldots, K$.
2. Given the data $D^l$ at the $l$-th interim look, $l = 1, 2, \ldots, L$,
   - (a) [Futility stopping] If the posterior probability that the hypothesis of arm $k$, $H_{1k}$, is true (i.e., $\lambda_k = 1$) is small, i.e.,
   $$Pr\{\lambda_k = 1 \mid D^l\} < P_{\text{futility}},$$
stop the accrual to the $k$-th arm for futility;
(b) Otherwise, continue to enroll patients until reaching the next interim analysis.

3. Once the maximum sample size is reached or all the arms have stopped, evaluate the efficacy for each arm based on all the observed data. If the posterior probability that that the hypothesis of arm $k$, $H_{1k}$, is true (i.e., $\lambda_k = 1$) is large, i.e.,

$$Pr\{\lambda_k = 1 \mid \mathcal{D}^{L+1}\} > \phi_k,$$

arm $k$ is declared efficacious and promising; otherwise, it is considered not promising.

Similar in BBHM (§11.3.1), Step 2 is optional. In U-Design, the probability threshold for futility interim analysis, $P_{futility}$, and for the final analysis, $\{\phi_k : k = 1, 2, \ldots, K\}$, are calibrated through simulations to achieve a prespecified type I error rate for each arm, under the null scenario. See the detailed calibration process in §11.3.1.

11.3.4.3 Discussion

MUCE is also used as a design for cohort expansion clinical trials. Finally, MUCE is a sophisticated method, the detail of which is in Lyu et al. (2020).
Part VI

Sample Size Calculation
12. Sample Size Calculation for Binary Outcome

In this Module, we implement the sample size calculation for binary endpoint, which include the following functions shown in Table 12.1.

Table 12.1: Function implementation in sample size calculation for binary endpoint.

<table>
<thead>
<tr>
<th>Number of Arms</th>
<th>Test Objectives</th>
<th>One- or/and Two-sided</th>
<th>Contents</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>Equality</td>
<td>One-sided &amp; Two-sided</td>
<td>Z-test</td>
<td>Section 12.1.1</td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>-</td>
<td>Z-test</td>
<td>Section 12.1.2</td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>-</td>
<td>Z-test</td>
<td>Section 12.1.3</td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>-</td>
<td>Z-test</td>
<td>Section 12.1.3</td>
</tr>
<tr>
<td></td>
<td>Agreement</td>
<td>One-sided &amp; Two-sided</td>
<td>Cohen’s Kappa</td>
<td>Section 12.1.4</td>
</tr>
<tr>
<td>Two (independent)</td>
<td>Equality</td>
<td>One-sided &amp; Two-sided</td>
<td>Z-test</td>
<td>Section 12.2.1</td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>-</td>
<td>Z-test</td>
<td>Section 12.2.2</td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>-</td>
<td>Z-test</td>
<td>Section 12.2.3</td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>-</td>
<td>Z-test</td>
<td>Section 12.2.3</td>
</tr>
<tr>
<td>Two (paired)</td>
<td>McNemar’s test</td>
<td>One-sided &amp; Two-sided</td>
<td></td>
<td>Section 12.3</td>
</tr>
</tbody>
</table>

12.1 Single arm

Let $x_i, i = 1, \ldots, n$ be the binary response observed from $i$th subject. In clinical research, $x_i$ could be the indicator for the response of tumor in cancer trials, i.e., $x_i = 1$ for responder or $x_i = 0$ for non-responder. It is assumed that $x_i$’s are i.i.d. with $P(x_i = 1) = p$, where $p$ is the true response
rate. Since \( p \) is unknown, it is usually estimated by

\[
\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\]

Also, let \( \epsilon = p - p_0 \) be the difference between the true response rate of a test drug (\( p \)) and a reference value (\( p_0 \)). In practice, it is of interest to test for equality (i.e., \( p = p_0 \)), non-inferiority (i.e., \( p - p_0 \) is greater than or equal to a pre-determined non-inferiority margin), superiority (i.e., \( p - p_0 \) is greater than a pre-determined superiority margin), and equivalence (i.e., the absolute difference between \( p \) and \( p_0 \) is within a difference of clinical importance). The following are details of sample size calculation with single arm.

**12.1.1 Test Objective: Equality**

**12.1.1.1 Methods**

- **Hypothesis:** To test whether there is a difference between the true response rate of the test drug and the reference value, the following hypotheses are usually considered,

  **(Two – sided)**

  \[
  H_0 : \epsilon = 0 \quad versus \quad H_1 : \epsilon \neq 0
  \]

  **(One – sided)**

  \[
  H_0 : \epsilon \leq 0 \quad versus \quad H_1 : \epsilon > 0
  \]

- **Formula:** Using the value of \( p \) to compute the standard deviation in \( z \)-test statistic, we can get sample size \( n \) from,

  **(Two – sided)**

  \[
  n = \frac{(z_{\alpha/2} + z_\beta)^2 p(1 - p)}{\epsilon^2}
  \]

  **(One – sided)**

  \[
  n = \frac{(z_\alpha + z_\beta)^2 p(1 - p)}{\epsilon^2}
  \]

  where \( z_\alpha \) is the upper \( \alpha \)th quantile of the standard normal distribution.

**12.1.1.2 Input and Output**

- **Input:**

  1. \( p_0 \): a reference value(response rate for the historical control)
2. $p$: true response rate of the test drug
3. $\alpha$: type I error rate
4. $\beta$: type II error rate (Power: $1 - \beta$)

- **Output**: sample size $n$

### 12.1.1.3 An Example (Single-arm Equality Two-sided Test)

Suppose that the response rate of the patient population under study after treatment by a test drug is expected to be around 50% (i.e., $p = 0.50$). At $\alpha = 0.05$, the required sample size for having an 80% power (i.e., $1 - \beta = 0.8$) for correctly detecting a difference between the post-treatment response rate and the reference value of 30% (i.e., $p_0 = 0.30$) can be obtained by the following steps,

- Select **SAMPLE SIZE**: Binary Outcome.
- Select **Number of Groups**: One, **Test Objective**: Equality and **1 or 2 Sided Test**: 2-Sided.
- Input $p_0$, $p$, $\alpha$ and $1 - \beta$.
- Click **Submit**.

Then the computed sample size is 50 using Z-test in this situation, shown in Figure 12.1.

### 12.1.2 Test Objective: Equivalence

#### 12.1.2.1 Methods

- **Hypothesis**: To establish equivalence, the following hypotheses are usually considered,

$$H_0 : |p - p_0| \geq \delta \quad versus \quad H_1 : |p - p_0| < \delta,$$

or

$$H_0 : |\epsilon| \geq \delta \quad versus \quad H_1 : |\epsilon| < \delta,$$

The proportion of the responses is concluded to be equivalent to the reference value of $p_0$ if the null hypothesis is rejected at a given significance level.

- **Formula**: Using the value of $p$ to compute the standard deviation in $z$-test statistic, we can get sample size $n$ from,

$$n = \frac{(z_\alpha + z_{\beta/2})^2 p(1 - p)}{(\delta - |\epsilon|)^2}$$
12.1.2.2 Input and Output

- **Input:**
  1. \( \delta (\delta > 0) \): equivalence margin
  2. \( p_0 \): a reference value
  3. \( p \): true response rate of the test drug
  4. \( \alpha \): type I error rate
  5. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- **Output:** sample size \( n \)

12.1.2.3 An Example (Single-arm Equivalence Test)

Assume that one brand name drug for a certain disease on the market has a responder rate of 60% (i.e., \( p_0 = 0.60 \)). It is believed that a 20% difference in responder rate is of no clinical significance.
12.1. Single arm
12.1.3. Test Objective: Non-Inferiority/Superiority

(i.e., $\delta = 0.2$). Hence, the investigator wants to show the study drug is equivalent to the market drug in terms of responder rate. At $\alpha = 0.05$, assuming that the true response rate is 60% (i.e., $p = 0.60$), the sample size required for achieving an 80% power can be obtained by the following steps,

- Select **SAMPLE SIZE: Binary Outcome**.

- Select **Number of Groups: One** and **Test Objective: Equivalence**.

- Input $\delta, p_0, p, \alpha$ and $1 - \beta$.

- Click **Submit**.

Then the computed sample size is 52 using Z-test in this situation, shown in Figure 12.2.

**Figure 12.2:** An Example (Single-arm Equivalence Test)
12.1.3 Test Objective: Non-Inferiority/Superiority

12.1.3.1 Methods

- **Hypothesis:** The problem of testing non-inferiority and superiority can be translated into the following hypotheses,
  
  \((\text{Noninferiority})\)
  \[
  H_0 : \epsilon \leq -\delta \quad \text{versus} \quad H_1 : \epsilon > -\delta
  \]

  \((\text{Superiority})\)
  \[
  H_0 : \epsilon \leq \delta \quad \text{versus} \quad H_1 : \epsilon > \delta
  \]

  where \(\delta (\delta > 0)\) is the superiority or non-inferiority margin.

- **Formula:** Using the value of \(p\) to compute the standard deviation in z-test statistic, we can get sample size \(n\) from,
  
  \((\text{Noninferiority})\)
  \[
  n = \frac{(z_\alpha + z_\beta)^2 p(1 - p)}{(\epsilon + \delta)^2}.
  \]

  \((\text{Superiority})\)
  \[
  n = \frac{(z_\alpha + z_\beta)^2 p(1 - p)}{(\epsilon - \delta)^2}.
  \]

12.1.3.2 Input and Output

- **Input:**
  1. \(\delta (\delta > 0)\): non-inferiority or superiority margin
  2. \(p_0\): a reference value
  3. \(p\): true response rate of the test drug
  4. \(\alpha\): type I error rate
  5. \(\beta\): type II error rate (Power: \(1 - \beta\))

- **Output:** sample size \(n\)

12.1.3.3 An Example (Single-arm Non-Inferiority Test)

For a certain disease, we wish to show that the majority of patients whose change after treatment by a test drug is at least as good as the reference value (30%) \((p_0 = 0.3)\). Also assume that a
difference of 10% in responder rate is considered of no clinical significance ($\delta = 0.1$). Assume the true response rate is 50% ($p = 0.5$). At $\alpha = 0.05$, the required sample size for having an 80% power (i.e., $1 - \beta = 0.8$) can be obtained by the following steps,

- Select **SAMPLE SIZE: Binary Outcome**.
- Select **Number of Groups: One** and **Test Objective: Non-Inferiority**.
- Input $\delta, p_0, p, \alpha$ and $1 - \beta$.
- Click **Submit**.

Then the computed sample size is 18 using Z-test in this situation, shown in Figure 12.3.

**Figure 12.3**: An Example (Single-arm Non-Inferiority Test)

12.1.4 Cohen’s Kappa

In some clinical trials, to check inter-rater reliability, independent sets of measurements are taken by more than one rater and the responses are checked for agreement. For a binary response, Cohens
Kappa test can be used to check inter-rater reliability. Conventionally, the kappa coefficient is used to express the degree of agreement between two raters when the same two raters rate each of a sample of \( n \) subjects independently. A simple example is given in the Table 12.2, where \( p_{ij} \) denotes the true proportion of the corresponding evaluations by Rater 1 and Rater 2 (e.g., \( p_{10} \) denotes that Rater 1 thinks it’s positive but Rater 2 thinks it’s negative), \( p_i. = p_{i1} + p_{i0} \) and \( p. j = p_{1j} + p_{0j} \).

### Table 12.2: Proportional Distribution by Two Rater

<table>
<thead>
<tr>
<th>Rater 1</th>
<th>positive</th>
<th>negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>( p_{11} )</td>
<td>( p_{01} )</td>
</tr>
<tr>
<td>negative</td>
<td>( p_{10} )</td>
<td>( p_{10} )</td>
</tr>
</tbody>
</table>

Kappa coefficient \( \kappa \) takes the form,

\[
\kappa = \frac{p_o - p_e}{1 - p_e}
\]

where \( p_o (p_o = p_{11} + p_{00}) \) is the proportion of rater pairs exhibiting agreement and \( p_e (p_e = p_{1.} p_{1.} + p_{0.} p_{0.}) \) is the proportion expected to exhibit agreement by chance alone. Thus "perfect agreement" would be indicated by \( \kappa = 1 \), and no agreement (other than that expected by chance) means that \( \kappa = 0 \).

**12.1.4.1 Methods**

- **Hypothesis:** The hypotheses of interest are

  (Two – sided)

  \[
  H_0 : \kappa = k_0 \text{ versus } H_1 : \kappa \neq k_1
  \]

  (One – sided)

  \[
  H_0 : \kappa = k_0 \text{ versus } H_1 : \kappa > k_1
  \]

- **Formula:** We can get sample size \( n \) from,

  (Two – sided)

  \[
  n = \left[ \frac{z_{\alpha/2} \sqrt{Q_0} + z_{\beta} \sqrt{Q_1}}{k_1 - k_0} \right]^2
  \]

  (One – sided)

  \[
  n = \left[ \frac{z_{\alpha} \sqrt{Q_0} + z_{\beta} \sqrt{Q_1}}{k_1 - k_0} \right]^2
  \]
where \( Q_0 (Q_1) \) can be calculated by using \( k_0 \) (\( k_1 \)) with

\[
Q_0(Q_1) = (1 - p_e)^{-4} \left\{ \sum_i p_{ii} [(1 - p_e) - (p_i + p_i)(1 - p_0)]^2 + (1 - p_0)^2 \sum_{i \neq j} p_{ij} (p_i + p_j)^2 - (p_0 p_e - 2 p_e + p_0)^2 \right\}
\]

Note that all of the values needed are uniquely determined by \( p_1, p_{.1}, k_0 \) and \( k_1 \). Specifically,

\[
p_0 = 1 - p_1,
\]
\[
p_{.0} = 1 - p_{.1}
\]
\[
p_e = p_{1.1} + p_{0.0}
\]
\[
p_o = \begin{cases} k_0 (1 - p_e) + p_e & \text{for } Q_0 \\ k_1 (1 - p_e) + p_e & \text{for } Q_1 \end{cases}
\]
\[
p_{00} = (p_0 - p_{1.} - p_{.0})/2
\]
\[
p_{11} = p_0 - p_{00}
\]
\[
p_{10} = p_{1.} - p_{11}
\]
\[
p_{01} = p_{.1} - p_{11}
\]

### 12.1.4.2 Input and Output

- **Input:**
  1. \( p_{1.} \): proportion that Rater 1 gives positive evaluation
  2. \( p_{.1} \): proportion that Rater 2 gives positive evaluation
  3. \( k_0 \): reference value of Kappa coefficient
  4. \( k_1 \): expected value of Kappa coefficient
  5. \( \alpha \): type I error rate
  6. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- **Output:** sample size \( n \)
12.1.4.3 An Example (Single-arm Cohen’s Kappa Test)

As an example, suppose two evaluation methods are asked to rate a group of cancer patients and to decide whether or not the status of each exhibits positive. We expect each method to identify 20% of patients to be positive ($p_1 = p_{.1} = 0.20$). Let $\kappa$ denote the level of agreement. The null hypothesis is $H_0: \kappa = 0.6$, but we expect Kappa coefficient is 0.9. At $\alpha = 0.05$, the required sample size for having an 80% power (i.e., $1 - \beta = 0.8$) can be obtained by the following steps,

- Select **SAMPLE SIZE: Binary Outcome**.
- Select **Number of Groups: One**, **Test Objective: Agreement(Cohen’s Kappa)** and **1 or 2 Sided Test: 2-Sided**.
- Input $p_{1.}, p_{.1}, k_0 = 0.6, k_1 = 0.9, \alpha$ and $1 - \beta$.
- Click **Submit**.

Then the computed sample size is 67 in this situation, shown in Figure 12.4.

12.2 Two arms (independent)

Let $x_{ij}$ be a binary response from the $j$th subject in the $i$th treatment group, $j = 1, \cdots, n_i, i = 1, 2$. For a fixed $i$, it is assumed that $x_{ij}$’s are i.i.d. with $P(x_{ij} = 1) = p_i$. In practice, $p_i$ is usually estimated by the observed proportion in the $i$th treatment group,

$$\hat{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}.$$

Let $\epsilon = p_t - p_c$ be the difference between the true mean response rates of a test drug ($p_t$) and a control ($p_c$). It is of interest to test for equality (i.e., $p_t = p_c$), non-inferiority (i.e., $p_t - p_c$ is greater than or equal to a pre-determined non-inferiority margin), superiority (i.e., $p_t - p_c$ is greater than a pre-determined superiority margin), and equivalence (i.e., the absolute difference between $p_t$ and $p_c$ is within a difference of clinical importance). The following are details of sample size calculation with two arms.
12.2. Two arms (independent)
12.2.1. Test Objective: Equality

12.2.1 Methods

- **Hypothesis:** To test whether there is a difference between the mean response rates of the test drug and the reference drug, the following hypotheses are usually considered,
  
  (Two - sided)
  
  \[ H_0 : \epsilon = 0 \quad versus \quad H_1 : \epsilon \neq 0 \]

  (One - sided)
  
  \[ H_0 : \epsilon \leq 0 \quad versus \quad H_1 : \epsilon > 0 \]

- **Formula:** We can get sample sizes \( n_t \) and \( n_c \) from
  
  (Two - sided)
  
  \[ n_c = \frac{(z_{\alpha/2} + z_{\beta})^2}{\epsilon^2} \left[ \frac{p_t(1-p_t)}{k} + p_c(1-p_c) \right] \]

  (One - sided)
  
  \[ n_c = \frac{(z_{\alpha} + z_{\beta})^2}{\epsilon^2} \left[ \frac{p_t(1-p_t)}{k} + p_c(1-p_c) \right] \]

  and \( n_t = kn_c \)

12.2.1.2 Input and Output

- **Input:**
  
  1. \( p_c \): true response rate of control treatment
  2. \( p_t \): true response rate of the test drug
  3. \( k \) \( (k = n_t/n_c) \): subject ratio of test control versus treatment
  4. \( \alpha \): type I error rate
  5. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- **Output:** sample sizes \( n_t \) and \( n_c \)

12.2.1.3 An Example (Two-arms (independent) Equality Two-sided Test)

In this example, suppose that a difference of \( \epsilon = 20\% \) in clinical response of cure is considered of clinically meaningful difference between the two agents for a certain disease. Assuming that the true cure rate for control treatment and the test drug are 65\% \( (p_c = 0.65 \) and \( p_t = p_c + \epsilon = 0.85 \)), respectively, at \( \alpha = 0.05 \), the sample sizes for having an 80\% power (i.e., \( 1 - \beta = 0.8 \)) with \( k = 1 \) (equal allocation) can be determined by the following steps,
Module 12. Sample Size Calculation for Binary Outcome

- Select SAMPLE SIZE: Binary Outcome.
- Select Number of Groups: Two, Test Objective: Equality and 1 or 2 Sided Test: 2-Sided.
- Input $p_c$, $p_t$, $k$, $\alpha$ and $1 - \beta$.
- Click Submit.

Then the computed sample sizes in this situation are shown in Figure 12.5.

12.2.2 Test Objective: Equivalence

12.2.2.1 Methods

- **Hypothesis:** To establish equivalence, the following hypothesis is usually considered,
  \[ H_0 : |\epsilon| \geq \delta \quad versus \quad H_1 : |\epsilon| < \delta \]

- **Formula:** We can get sample sizes $n_t$ and $n_c$ from
  \[
  n_c = \left( z_\alpha + z_{\beta/2} \right)^2 \frac{p_t(1-p_t)}{k} + p_c(1-p_c) \]
  and
  \[
  n_t = kn_c.
  \]

12.2.2.2 Input and Output

- **Input:**
  1. $\delta$ ($\delta > 0$): equivalence margin
  2. $p_c$: true response rate of control treatment
  3. $p_t$: true response rate of the test drug
  4. $k$ ($k = n_t/n_c$): subject ratio of test control versus treatment
  5. $\alpha$: type I error rate
  6. $\beta$: type II error rate (Power: $1 - \beta$)

- **Output:** $n_t$ and $n_c$

12.2.2.3 An Example (Two-arms (independent) Equivalence Test)

For establishment of equivalence, suppose the true cure rate for the two agents are 75% ($p_c = 0.75$) and 80% ($p_t = 0.80$) and the equivalence limit is 20% (i.e., $\delta = 0.20$). At $\alpha = 0.05$, the sample
12.2. Two arms (independent)

12.2.3. Test Objective: Non-Inferiority/Superiority

sizes for having an 80% power (i.e., $1 - \beta = 0.8$) with $k = 1$ (equal allocation) can be determined by the following steps,

- Select **SAMPLE SIZE: Binary Outcome**.
- Select **Number of Groups: Two** and **Test Objective: Equivalence**.
- Input $\delta, p_c, p_t, k, \alpha$ and $1 - \beta$.
- Click **Submit**.

Then the computed sample sizes in this situation are shown in Figure 12.6.

### 12.2.3 Test Objective: Non-Inferiority/Superiority

#### 12.2.3.1 Methods

- **Hypothesis**: The problem of testing non-inferiority and superiority can be translated into the following hypotheses,

  (**Noninferiority**)  
  $$H_0: \epsilon \leq -\delta \text{ versus } H_1: \epsilon > -\delta$$

  (**Superiority**)  
  $$H_0: \epsilon \leq \delta \text{ versus } H_1: \epsilon > \delta$$

  where $\delta$ ($\delta > 0$) is the superiority or non-inferiority margin.

- **Formula**: We can get sample sizes $n_t$ and $n_c$ from

  (**Noninferiority**)  
  $$n_c = \frac{(z_\alpha + z_\beta)^2}{(\epsilon + \delta)^2} \left[ \frac{p_t(1 - p_t)}{k} + p_c(1 - p_c) \right]$$

  (**Superiority**)  
  $$n_c = \frac{(z_\alpha + z_\beta)^2}{(\epsilon - \delta)^2} \left[ \frac{p_t(1 - p_t)}{k} + p_c(1 - p_c) \right]$$

#### 12.2.3.2 Input and Output

- **Input**:
  1. $\delta$ ($\delta > 0$): non-inferiority or superiority margin
2. \( p_0 \): a reference value
3. \( p \): true response rate of the test drug
4. \( \alpha \): type I error rate
5. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- **Output**: sample sizes \( n_t \) and \( n_c \)

### 12.2.3.3 An Example (Two-arms (independent) Non-Inferiority Test)

Now, suppose it is of interest to establish non-inferiority of the test drug as compared to the active control agent. Similarly, we consider the difference less than 10% is of no clinical importance. Thus, the non-inferiority margin is chosen to be 10% (i.e., \( \delta = 0.10 \)). Also, suppose the true mean cure rates of the treatment agents and the active control are 85% and 65% (i.e., \( p_t = 0.85 \) and \( p_c = 0.65 \)), respectively. Then, at \( \alpha = 0.05 \), the sample size for having an 80% power (i.e., \( 1 - \beta = 0.8 \)) with \( k = 1 \) (equal allocation) can be determined by the following steps,

- Select **SAMPLE SIZE**: Binary Outcome.
- Select **Number of Groups**: One and **Test Objective**: Non-Inferiority.
- Input \( \delta, p_c, p_t, k, \alpha \) and \( 1 - \beta \).
- Click **Submit**.

Then the computed sample sizes in this situation are shown in Figure 12.7.

### 12.3 Two arms (paired): McNemar’s Test

For a given laboratory test, test results are usually summarized as either normal or abnormal. Let \( x_{ij} \) denote the binary response from the \( i \)th \((i = 1, 2, \ldots, n)\) subject in the \( j \)th treatment where \( j = 1 \) denotes pre-treatment and \( j = 2 \) post-treatment, and \( x_{ij} = 1 \) denotes that the response is normal and \( x_{ij} = 0 \) abnormal. The test results can be summarized in Table 12.3, where \( n_{ij}, i, j = 1, 0 \) are defined by as follows,

\[
\begin{align*}
    n_{11} &= \sum_{i=1}^{n} x_{i1}x_{i2} \\
    n_{10} &= \sum_{i=1}^{n} x_{i1}(1 - x_{i2})
\end{align*}
\]
12.3. Two arms (paired): McNemar’s Test

12.3.1. Methods

Table 12.3: Test Results of Two Arms Paired

<table>
<thead>
<tr>
<th>Pre-treatment</th>
<th>Post-treatment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>normal</td>
<td>$n_{11}$</td>
</tr>
<tr>
<td>abnormal</td>
<td>abnormal</td>
<td>$n_{10}$</td>
</tr>
<tr>
<td></td>
<td>$n_{01}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n_{00}$</td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
  n_{01} &= \sum_{i=1}^{n} (1 - x_{i1}) x_{i2} \\
  n_{00} &= \sum_{i=1}^{n} (1 - x_{i1})(1 - x_{i2})
\end{align*}
\]

Define,

\[
\begin{align*}
  p_{11} &= n_{11}/n \\
  p_{10} &= n_{10}/n \\
  p_{01} &= n_{01}/n \\
  p_{00} &= n_{00}/n \\
  p_{1.} &= p_{11} + p_{10} \\
  p_{-1} &= p_{11} + p_{01}
\end{align*}
\]

12.3.1 Methods

- **Hypothesis**: It is of interest to test whether there is a categorical shift after treatment. A categorical shift is defined as either a shift from 0 (abnormal) in pre-treatment to 1 (normal) in post-treatment or a shift from 1 (normal) in pre-treatment to 0 (abnormal) in post-treatment. Thus, the hypothesis of interest is

  (Two – sided)

  \[ H_0 : p_{1.} = p_{-1} \quad \text{versus} \quad H_1 : p_{1.} \neq p_{-1} \]

  (One – sided)

  \[ H_0 : p_{1.} = p_{-1} \quad \text{versus} \quad H_1 : p_{1.} > p_{-1} \]

  which is equivalent to

  (Two – sided)

  \[ H_0 : p_{10} = p_{01} \quad \text{versus} \quad H_1 : p_{10} \neq p_{01} \]
Module 12. Sample Size Calculation for Binary Outcome

(One – sided)

\[ H_0 : p_{10} = p_{01} \quad versus \quad H_1 : p_{10} > p_{01} \]

**Formula:** We can get sample size \( n \) from

(Two – sided)

\[ n = \left[ z_{\alpha/2} \sqrt{p_{10} + p_{01}} + z_{\beta} \sqrt{p_{10} + p_{01} - (p_{10} - p_{01})^2} \right]^2 \frac{1}{(p_{10} - p_{01})^2} \]

(One – sided)

\[ n = \left[ z_{\alpha} \sqrt{p_{10} + p_{01}} + z_{\beta} \sqrt{p_{10} + p_{01} - (p_{10} - p_{01})^2} \right]^2 \frac{1}{(p_{10} - p_{01})^2} \]

12.3.2 Input and Output

- **Input:**
  1. \( p_{10} \): probability of shifting from normal to abnormal
  2. \( p_{01} \): probability of shifting from abnormal to normal
  3. \( \alpha \): type I error rate
  4. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- **Output:** sample size \( n \)

12.3.3 An Example (Two-arms (paired) McNemar’s Test)

Consider a study, it is expected that about 50% \( (p_{10} = 0.50) \) of patients will shift from 1 (abnormal pre-treatment) to 0 (normal post-treatment) and 20% \( (p_{01} = 0.20) \) of patients will shift from 0 (normal pre-treatment) to 1 (abnormal post-treatment).

The investigator would like to select a sample size such that there is an 80% \( (1 - \beta = 0.80) \) power for detecting such a difference if it truly exists at the 5% \( (\alpha = 0.05) \) level of significance.

The required sample size can be obtained as follows:

- Select **SAMPLE SIZE:** Binary Outcome.
- Select **Number of Groups:** One and **Test Objective:** Non-Inferiority.
- Input \( \delta, p_c, p_t, k, \alpha \) and \( 1 - \beta \).
- Click **Submit**.

Then the computed sample size in this situation is shown in Figure 12.8.
12.3. Two arms (paired): McNemar’s Test

12.3.3. An Example (Two-arms (paired) McNemar’s Test)

**Figure 12.4:** An Example (Single-arm Cohen’s Kappa Test)

Sample Size Calculation - Binary Outcome

Superiority Test for One-Sample Mean

Cohen’s Kappa

In a two-sided test for agreement using Kappa’s Coefficient, at the significance level of 0.05, a sample size of 67 is needed to achieve 80% power when the probability that Rater 1 will give positive evaluation is 0.2 and the probability that Rater 2 will give positive evaluation is 0.2.
**Figure 12.5:** An Example (Two-arms (independent) Equality Two-sided Test)

Sample Size Calculation - Binary Outcome

Two-Sample Two-Sided Test for Equal Means

**Z-test**

In a two-sided z test for two-sample mean, at the significance level of 0.05, 70 subjects for the treatment group and 70 subjects for the control group are needed to achieve 80% when the response rate for control is 0.65 and the response rate for the test drug is 0.85.
12.3. Two arms (paired): McNemar’s Test

12.3.3. An Example (Two-arms (paired) McNemar’s Test)

**Figure 12.6:** An Example (Two-arms (independent) Equivalence Test)

Sample Size Calculation - Binary Outcome

Two-Sample Equivalence Test

- **Z-test**
  - At the significance level of 0.05, with an equivalence limit of 0.2, 133 subjects for the treatment group and 133 subjects for the control group are needed to achieve 80% power when the response rate for control is 0.75 and the response rate for the test drug is 0.8.
Figure 12.7: An Example (Two-arms (independent) Non-Inferiority Test)

At the significance level of 0.05, with an non-inferiority margin of 0.1, 25 subjects for the treatment group and 26 subjects for the control group are needed to achieve 80% power when the response rate for control is 0.65 and the response rate for the test drug is 0.85.
Figure 12.8: An Example (Two-arms (paired) McNemar’s Test)
13. Sample Size Calculation for Continuous Outcome

In this module, we implement the sample size calculation for continuous endpoint, which include the following functions shown in Table 13.1.

Table 13.1: Function implementation in sample size calculation for binary endpoint.

<table>
<thead>
<tr>
<th>Number of Arms</th>
<th>Test Objectives</th>
<th>One-or/and Two-sided</th>
<th>Contents</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>Equality</td>
<td>One-sided &amp; Two-sided</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.1.1</td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>-</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.1.3</td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>-</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.1.2</td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>-</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.1.2</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>One-sided &amp; Two-sided</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.1.4</td>
</tr>
<tr>
<td>Two (independent)</td>
<td>Equality</td>
<td>One-sided &amp; Two-sided</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.2.1</td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>-</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.2.2</td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>-</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.2.3</td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>-</td>
<td>Z-test &amp; T-test</td>
<td>Section 13.2.3</td>
</tr>
<tr>
<td>Two (paired)</td>
<td>Paired</td>
<td>One-sided &amp; Two-sided</td>
<td>T-test</td>
<td>Section 13.3</td>
</tr>
<tr>
<td>Multiple</td>
<td>ANOVA</td>
<td>-</td>
<td>F-test</td>
<td>Section 13.4</td>
</tr>
</tbody>
</table>

13.1 Single arm

To compare a new drug to a placebo control, one single-sample study will be conducted. This single sample will consist of observations from a single treatment using the new drug when the mean is
13.1. Single arm

13.1.1 Test Objective: Equality

to be compared to a specified constant, the reference response. Let $\epsilon = \mu_t - \mu_c$ be the difference between the expected mean response ($\mu_t$) of the new drug and a reference response value ($\mu_c$) from the control. The main reference for this section is Chow et al. (2017).

13.1.1 Test Objective: Equality

13.1.1.1 Methods

To test whether there is a difference between the mean response of the test drug and the reference value, the following hypotheses and calculation formulas are usually considered,

- **Hypothesis:**

  - *(Two - sided)* if there is a difference between $\mu_t$ and $\mu_c$,

    $$H_0 : \epsilon = 0 \quad versus \quad H_1 : \epsilon \neq 0$$

  - *(One - sided)* if there is a positive difference between $\mu_t$ and $\mu_c$, that is $\mu_t > \mu_c$,

    $$H_0 : \epsilon \leq 0 \quad versus \quad H_1 : \epsilon > 0$$

- **Formula:**

  - for **T-test**, we search for sample size $n$ that satisfies the following conditions, *(Two - sided)*

    $$T_{n-1} \left\{ t_{\alpha/2, n-1} \left| \frac{\sqrt{n\epsilon^2}}{\sigma} \right. \right\} - T_{n-1} \left\{ -t_{\alpha/2, n-1} \left| \frac{\sqrt{n\epsilon^2}}{\sigma} \right. \right\} = \beta$$

    *(One - sided)*

    $$T_{n-1} \left\{ t_{\alpha, n-1} \left| \frac{\sqrt{n\epsilon^2}}{\sigma} \right. \right\} = \beta$$

  - for **Z-test**, we can get sample size $n$ from, *(Two - sided)*

    $$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\epsilon^2}$$

    *(One - sided)*

    $$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{\epsilon^2}$$
Module 13. Sample Size Calculation for Continuous Outcome

13.1.1.2 Input and Output

- **Input:**
  1. $\epsilon$: difference between the true mean response of a test drug ($\mu_t$) and a reference value ($\mu_c$)
  2. $\alpha$: type I error rate
  3. $\beta$: type II error rate (Power: $1 - \beta$)
  4. $\sigma$: standard deviation (we assume standard deviation is known when $z$-test and unknown when $t$-test) and $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

- **Output:** sample size $n$

13.1.1.3 An Example (Single-arm Equality Two-sided Test)

Consider an example concerning a study of osteoporosis (or decreased bone mass). Usually, the measure of bone density is SD.

Suppose that the mean bone density before the treatment is 1.5 SD ($\mu_c = 1.5$ SD) and after treatment is expected to be 2.0 SD ($\mu_t = 2$ SD) with the standard deviation ($\sigma = 1$). At $\alpha = 0.05$, the required sample size for having an 80% power ($1 - \beta = 0.8$) for correctly detecting a difference of $\epsilon = 0.5$ SD change from pre-treatment to post-treatment can be obtained by the following steps,

- Select **SAMPLE SIZE: Continuous Outcome**.
- Select **Number of Groups: One**, **Test Objective: Equality** and **1 or 2 Sided Test: 2-Sided**.
- Input $\mu_t, \mu_c, \sigma, \alpha$ and $1 - \beta$.
- Click **Submit**.

Then the computed sample sizes are 34 using $t$-test and 32 using $Z$-test in this situation, shown in Figure 13.1.

13.1.2 Test Objective: Non-Inferiority/Superiority

13.1.2.1 Methods

The problem of testing non-inferiority and superiority can be explained by the following hypotheses,

- **Hypothesis:**
13.1. Single arm

13.1.2. Test Objective: Non-Inferiority/Superiority

Figure 13.1: An Example (Single-arm Equality Two-sided Test)

- (Non–in inferiority) if the new drug $\mu_t$ is not much worse than the placebo control $\mu_c$. In other words, $\epsilon = \mu_t - \mu_c$ is not too small,

$$H_0 : \epsilon \leq -\delta \quad \text{versus} \quad H_1 : \epsilon > -\delta$$

- (Superiority) if the new drug $\mu_t$ is much better than the placebo control $\mu_c$. In other words, $\epsilon = \mu_t - \mu_c$ is big enough,

$$H_0 : \epsilon \leq \delta \quad \text{versus} \quad H_1 : \epsilon > \delta$$

where $\delta$ ($\delta > 0$) is the non-inferiority or superiority margin.

- Formula:

  - for T-test, we search for a $n$ that satisfies
Module 13. Sample Size Calculation for Continuous Outcome

(Noninferiority)
\[ T_{n-1} \left\{ t_{\alpha,n-1} \frac{\sqrt{n}(\epsilon + \delta)}{\sigma} \right\} = \beta \]

(Superiority)
\[ T_{n-1} \left\{ t_{\alpha,n-1} \frac{\sqrt{n}(\epsilon - \delta)}{\sigma} \right\} = \beta \]

– for Z-test, we can get sample size \( n \) from

(Noninferiority)
\[ n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{(\epsilon + \delta)^2} \]

(Superiority)
\[ n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{(\epsilon - \delta)^2} \]

13.1.2.2 Input and Output

- **Input:**
  1. \( \delta \): superiority or non-inferiority margin
  2. \( \epsilon \): difference between the true mean response of a test drug (\( \mu_t \)) and a reference value (\( \mu_c \))
  3. \( \alpha \): type I error rate
  4. \( \beta \): type II error rate (Power: \( 1 - \beta \))
  5. \( \sigma \): standard deviation (we assume standard deviation is known when z-test and unknown when t-test) and \( \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \)

- **Output:** sample size \( n \)

13.1.2.3 An Example (Single-arm Non-inferiority Test)

In the study of osteoporosis, we wish to show that the mean bone density post-treatment is no less than pre-treatment by a clinically meaningful difference \( \delta = 0.5 \) SD. We know mean bone density pre-treatment is 1.5 (\( \mu_c = 1.5 \)). Suppose the expected mean bone density post-treatment is 2.0 (\( \mu_t = 2.0 \)) with standard deviation of 1 (\( \sigma = 1 \)). At \( \alpha = 0.025 \), the required sample size for having an 80% power (\( 1 - \beta = 0.8 \)) can be obtained by the following steps,

- Select **SAMPLE SIZE: Continuous Outcome.**
13.1. Single arm

13.1.3. Test Objective: Equivalence

- Select **Number of Groups: One** and **Test Objective: Non-inferiority**.

- Input $\delta, \mu_c, \mu_t, \sigma, \alpha$ and $1 - \beta$.

- Click **Submit**.

Then the computed sample sizes are 10 using T-test and 8 using Z-test in this situation, shown in Figure 13.2.

**Figure 13.2:** An Example (Single-arm Non-inferiority Test)
Module 13. Sample Size Calculation for Continuous Outcome

13.1.3 Test Objective: Equivalence

13.1.3.1 Methods

The objective is to test how close the treatment effect of the test drug is to a gold standard on average. The following hypothesis will be considered,

\[ H_0 : |\epsilon| \geq \delta \text{ versus } H_1 : |\epsilon| < \delta. \]

- For T-test, we search for sample size \( n \) that satisfies
  \[ T_{n-1} \left\{ t_{\alpha,n-1} \sqrt{n(\delta - |\epsilon|)} \right\} = \frac{\beta}{2} \]

- For Z-test, we can get sample size \( n \) from,
  \[ n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\delta - |\epsilon|)^2} \]

13.1.3.2 Input and Output

- Input:
  1. \( \delta \): equivalence margin; \( \delta > 0 \)
  2. \( \epsilon \): difference between the true mean response of a test drug (\( \mu_t \)) and a reference value (\( \mu_c \))
  3. \( \sigma \): standard deviation (we assume standard deviation is known when z-test and unknown when t-test) and \( \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_1 - \bar{x})^2 \)
  4. \( \alpha \): type I error rate
  5. \( 1 - \beta \): power (\( \beta \) is type II error rate)

- Output: sample size \( n \)

13.1.3.3 An Example (Single-arm Equivalence Test)

Consider an example concerning the effect of a test drug on body weight change in terms of body mass index (BMI) before and after the treatment.

Suppose clinicians consider that a less than 5% change in BMI from baseline (pre-treatment) to endpoint (post-treatment) is not a safety concern for the indication of the disease under study. Thus, we consider \( \delta = 0.05 \) as the equivalence margin. The objective is then to demonstrate safety by
13.1. Single arm

13.1.4. Test Objective: Correlation

Testing equivalence in mean BMI between pre-treatment and post-treatment of the test drug. Assume the true BMI before and after the treatment are both 0.2 ($\mu_c = \mu_t = 0.2$) and the difference of them is 0 ($\epsilon = 0$) and the standard deviation is 10% ($\sigma = 0.1$), with $\alpha = 0.05$, the sample size required for achieving an 80% power ($1 - \beta = 0.8$) can be obtained by the following steps,

- Select SAMPLE SIZE: Continuous Outcome.
- Select Number of Groups: One and Test Objective: Equivalence.
- Input $\delta, \mu_c, \mu_t, \sigma, \alpha$ and $1 - \beta$.
- Click Submit.

Then the computed sample sizes are 36 using T-test and 35 using Z-test in this situation, shown in Figure 13.3.

**Figure 13.3:** An Example (Single-arm Equivalence Test)
13.1.4 Test Objective: Correlation

This subsection introduces the single-arm correlation test. The correlation coefficient \( \rho \) is calculated as

\[
\rho = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}},
\]

indicating that the relationship of only two variables is being examined, e.g. the relationship of patient’s age \( (X) \) and treatment effect of a certain drug \( (Y) \). The main reference for this subsection is Zar (2010).

13.1.4.1 Methods

To test whether there is a correlation between two variables, the following hypotheses and calculation formulas are usually considered,

- **Hypothesis:**
  - *(Two - sided)* if there is a correlation between the two variables,
    \[
    H_0 : \rho = 0 \quad versus \quad H_1 : \rho = r
    \]
    where \( r \neq 0 \).
  - *(One - sided)* if there is a positive correlation between the two variables,
    \[
    H_0 : \rho = 0 \quad versus \quad H_1 : \rho > r
    \]
    where \( r > 0 \).

- **Formula:** We can use both t-test and z-test to calculate the sample size for the hypothesis. Both tests use Fishers Transformation, denoted as \( C(r) = 0.5 \log(\frac{1+r}{1-r}) \).

  - For **Z-test**: Given a sample correlation \( r \) based on \( n \) observations that are from a population with true correlation parameter \( \rho \), \( C(r) \) follows a normal distribution with mean \( C(\rho) \) and variance \( 1/\sqrt{n - 3} \).

    \[
    C(r) \sim N(C(\rho), 1/\sqrt{n - 3})
    \]

    Thus, under \( H_0 \), \( \sqrt{n - 3}C(r) \sim N(0, 1) \) since \( C(\rho) = 0.5 \log(1) = 0 \). The sample sizes required to achieve the power \( 1 - \beta \) and control type I error rate at \( \alpha \) are as follows:
13.1. Single arm

13.1.4. Test Objective: Correlation

\( (Two \ - \ sided) \)

\[ n = \left( \frac{z_{\alpha/2} + z_{\beta}}{C(r)} \right)^2 + 3 \]

\( (One \ - \ sided) \)

\[ n = \left( \frac{z_{\alpha} + z_{\beta}}{C(r)} \right)^2 + 3 \]

- For T-test: The t-test for significance of \( r \) is given by

\[ t = \frac{r \sqrt{n - 2}}{\sqrt{1 - r^2}}. \]

If we find the critical \( t \) value, denoted as \( t_c \), above which we will reject \( H_0 \), then we can get \( r_c \).

\[ r_c = \sqrt{\frac{t_c^2}{t_c^2 + n - 2}} \]

The sample size calculation involves the transformation proposed by Pearson and Hartley (1996):

\[ C_r = C(r) + \frac{r}{2(n - 1)}, \]

\[ C_{r_c} = C(r_c). \]

The sample size required can be obtained by solving the following equations iteratively.

\( (Two \ - \ sided) \)

\[ 1 - \beta = \Phi\{(C_r - C_{r_c})\sqrt{n - 3}\} + \Phi\{(-C_r - C_{r_c})\sqrt{n - 3}\} \]

\( (One \ - \ sided) \)

\[ 1 - \beta = \Phi\{(C_r - C_{r_c})\sqrt{n - 3}\} \]

13.1.4.2 Input and Output

- **Input:**
  1. \( r \): correlation coefficient under alternative hypothesis, or expected correlation coefficient
  2. \( \alpha \): type I error rate
  3. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- **Output:** sample size \( n \)
13.1.4.3 **An Example (Single-arm Correlation Test)**

Consider a situation where we want to test whether the treatment effect of a certain new drug is associated with the patient age. It’s hoped that the correlation coefficient between the treatment effect of this new drug and the patient age is 0.3 ($r = 0.3$). And we want the design with type I error rate of 0.05 ($\alpha = 0.05$) and power of 90% ($1 - \beta = 0.9$). The sample sizes can be obtained by the following steps,

- Select **SAMPLE SIZE: Continuous Outcome**.
- Select **Number of Groups: One, Test Objective: Correlation** and **1 or 2 Sided Test: 2-Sided**.
- Input $r$, $\alpha$ and $1 - \beta$.
- Click **Submit**.

This will calculate the sample sizes for this design and the output is shown in the right panel. The computed sample sizes are 112 using T-test and 113 using Z-test in this situation, shown in Figure 13.4.

13.2 **Two arms (independent)**

To compare a new drug to a standard treatment, one two-samples study will be conducted. These two samples will consist of observations from the treatment using this new drug and this standard treatment. Let $\epsilon = \mu_t - \mu_c$ be the difference between the expected mean response of this new drug ($\mu_t$) and this standard treatment ($\mu_c$). In practice, it may be desirable to have an unequal treatment allocation, i.e., $n_c/n_t = k$ for some $k$, where $n_t$ and $n_c$ denote sample sizes for treatment and control respectively. Note that $k = 1/2$ indicates a 2 to 1 test-control allocation, whereas $k = 2$ indicates a 1 to 2 test-control allocation.

13.2.1 **Test Objective: Equality**

13.2.1.1 **Methods**

To test whether there is a difference between the mean response of the test drug and the reference value, the following hypotheses and calculation formulas are usually considered,

- **Hypothesis:**
13.2. Two arms (independent)

13.2.1. Test Objective: Equality

Figure 13.4: An Example (Single-arm Correlation Test)

- (Two – sided) if there is a difference between \( \mu_t \) and \( \mu_c \),

\[
H_0 : \epsilon = 0 \quad \text{versus} \quad H_1 : \epsilon \neq 0
\]

- (One – sided) if there is a positive difference between \( \mu_t \) and \( \mu_c \), that is \( \mu_t > \mu_c \), or \( \epsilon > 0 \),

\[
H_0 : \epsilon \leq 0 \quad \text{versus} \quad H_1 : \epsilon > 0
\]

* Formula:

- for T-test, we search for \( n_t \) that satisfies

(Two – sided)

\[
T_{(1+k)n_t-2} \left\{ t_{\alpha/2,(1+k)n_t-2} \frac{\sqrt{n_t\epsilon^2}}{\sigma \sqrt{1 + 1/k}} \right\} - T_{(1+k)n_t-2} \left\{ -t_{\alpha/2,(1+k)n_t-2} \frac{\sqrt{n_t\epsilon^2}}{\sigma \sqrt{1 + 1/k}} \right\} = \beta
\]

(One – sided)

\[
T_{(1+k)n_t-2} \left\{ t_{\alpha,(1+k)n_t-2} \frac{\sqrt{n_t\epsilon^2}}{\sigma \sqrt{1 + 1/k}} \right\} = \beta
\]

Your trial designs anywhere, anytime.

299
Module 13. Sample Size Calculation for Continuous Outcome

and $n_c = kn_t$.

- for Z-test, we can get sample sizes $n_t$ and $n_c$ from,

\[ n_t = \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2 (1 + 1/k)}{\epsilon^2} \]

\[ n_c = \frac{(z_\alpha + z_\beta)^2 \sigma^2 (1 + 1/k)}{\epsilon^2} \]

and $n_c = kn_t$.

13.2.1.2 Input and Output

- **Input:**

1. $\epsilon = \mu_t - \mu_c$: the expected mean difference between a test drug ($\mu_t$) and a standard treatment ($\mu_c$)
2. $k = n_c/n_t$: treatment allocation ratio
3. $\alpha$: type I error rate
4. $\beta$: type II error rate (Power: $1 - \beta$)
5. $\sigma$: variance. Assume that variance is known when z-test and unknown t-test, we often use the pooled variance to estimate it.

\[ \hat{\sigma}^2 = \frac{1}{n_c + n_t - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) \]

- **Output:**

1. $n_t$: sample size of treatment group
2. $n_c$: sample size of control group

13.2.1.3 An Example (Two-arms (Independent) Equality Two-sided Test)

Consider a pharmaceutical company that is interested in conducting a clinical trial to compare two cholesterol lowering agents through a parallel design. The primary efficacy parameter is the low density lipoprotein (LDL), because most of the cholesterol is bound to LDLs. In what follows, we will consider the situation where the intended trial is for testing equality of mean responses in LDL.

In this example, suppose a difference of 5% ($\epsilon = \mu_t - \mu_c = 0.05$) in percent change of LDL is considered of clinically meaningful difference. Assuming that the standard deviation is 10%
13.2. Two arms (independent)

13.2.2. Test Objective: Equivalence

\(\sigma = 10\%\), with \(\alpha = 0.05\), the sample sizes required for achieving an 80% power \((1 - \beta = 0.8)\) can be obtained by the following steps,

- **Select** SAMPLE SIZE: Continuous Outcome.

- **Select** Number of Groups: Two (independent), Test Objective: Equality and 1 or 2 Sided Test: 2-Sided.

- **Input** \(\epsilon = \mu_t - \mu_c, \sigma, k, \alpha\) and \(1 - \beta\).

- **Click** Submit.

Then the computed sample sizes in this situation are shown in Figure 13.5.

**Figure 13.5:** An Example (Two-arms (Independent) Equality Two-sided Test)
13.2.2 Test Objective: Equivalence

13.2.2.1 Methods

The objective is to test how close the treatment effect of the test drug and the standard treatment are. The following hypothesis will be considered,

\[ H_0 : |\epsilon| \geq \delta \quad versus \quad H_1 : |\epsilon| < \delta \]

- For T-test, we search for \( n_t \) that satisfies

\[
T_{(1+k)n_c-2} \left\{ t_{\alpha,(1+k)n_c-2} \sqrt{\frac{\sigma}{\left( 1 + \frac{1}{k} \right)}} \frac{\sqrt{n_t} (\delta - |\epsilon|)}{\sigma \sqrt{1 + 1/k}} \right\} = \frac{\beta}{2}
\]

- For Z-test, we can get sample sizes \( n_t \) and \( n_c \) from

\[
n_t = \frac{(z_\alpha + z_{\beta/2})^2 \sigma^2 (1 + 1/k)}{(\delta - |\epsilon|)^2} \quad \text{and} \quad n_c = kn_t
\]

13.2.2.2 Input and Output

- Input:
  1. \( \delta \): equivalence margin
  2. \( \epsilon = \mu_t - \mu_c \): the true mean difference between a test drug (\( \mu_t \)) and a standard treatment (\( \mu_c \))
  3. \( k = n_c/n_t \): treatment allocation ratio
  4. \( \sigma \): variance (we assume variance is known when z-test and unknown t-test)
  5. \( \alpha \): type I error rate
  6. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- Output:
  1. \( n_t \): sample size of treatment group
  2. \( n_c \): sample size of control group

13.2.2.3 An Example (Two-arms (Independent) Equivalence Test)

Consider a pharmaceutical company that is interested in conducting a clinical trial to compare two cholesterol lowering agents through a parallel design. The primary efficacy parameter is the low...
density lipoprotein (LDL), because most of the cholesterol is bound to LDLs. In what follows, we will consider the situation where the intended trial is testing for therapeutic equivalence.

For establishment of equivalence, suppose the true mean difference is 1% ($\epsilon = 0.01$) and the equivalence limit is 5% ($\delta = 0.05$). Assuming that the standard deviation is 10% ($\sigma = 10\%$), with $\alpha = 0.05$, the sample sizes required for achieving an 90% power ($1 - \beta = 0.9$) can be obtained by the following steps,

- Select **SAMPLE SIZE**: Continuous Outcome.
- Select **Number of Groups**: Two (independent) and **Test Objective**: Equivalence.
- Input $\delta$, $\epsilon = \mu_t - \mu_c$, $\sigma$, $k$, $\alpha$ and $1 - \beta$.
- Click **Submit**.

Then the computed sample sizes in this situation are shown in Figure 13.6.

**Figure 13.6: An Example (Two-arms (Independent) Equivalence Test)**
13.2.3 Test Objective: Non-Inferiority/Superiority

13.2.3.1 Methods

The problem of testing non-inferiority and superiority can be explained by the following hypotheses,

- **Hypothesis:**
  - *(Non − inferiority)* The objective is to confirm that the new drug $\mu_t$ is not much worse than the standard treatment $\mu_c$. In other words, $\epsilon = \mu_t - \mu_c$ is not too small,

  \[ H_0 : \epsilon \leq -\delta \ \text{versus} \ \ H_1 : \epsilon > -\delta \]

  - *(Superiority)* The objective is to confirm that the new drug $\mu_t$ is much better than the standard treatment $\mu_c$. In other words, $\epsilon = \mu_t - \mu_c$ is big enough,

  \[ H_0 : \epsilon \leq \delta \ \text{versus} \ \ H_1 : \epsilon > \delta \]

  where $\delta$ ($\delta > 0$) is the superiority or non-inferiority margin.

- **Formula:**
  - For **T-test**, we search for $n_t$ that satisfies
    - *(Non − inferiority)*
    \[ T_{(1+k)n_t-2} \left\{ t_{\alpha,(1+k)n_t-2} \left| \frac{\sqrt{\mu_t(\epsilon + \delta)}}{\sigma \sqrt{1 + 1/k}} \right. \right\} = \beta \]
    
    *(Superiority)*
    \[ T_{(1+k)n_t-2} \left\{ t_{\alpha,(1+k)n_t-2} \left| \frac{\sqrt{\mu_t(\epsilon - \delta)}}{\sigma \sqrt{1 + 1/k}} \right. \right\} = \beta \]
  
  and $n_c = kn_t$.
  
  - For **Z-test**, we can get sample sizes $n_t$ and $n_c$ from
    - *(Non − inferiority)*
    \[ n_t = \frac{(z_\alpha + z_\beta)^2 \sigma^2 (1 + 1/k)}{(\epsilon + \delta)^2} \]
    
    *(Superiority)*
    \[ n_t = \frac{(z_\alpha + z_\beta)^2 \sigma^2 (1 + 1/k)}{(\epsilon - \delta)^2} \]
  
  and $n_c = kn_t$. 

Your trial designs anywhere, anytime.
13.2. Two arms (independent)

13.3.1. Methods

13.2.3.2 Input and Output

• Input:

1. $\delta$: superiority or non-inferiority margin
2. $\epsilon = \mu_t - \mu_c$: the true mean difference between a test drug ($\mu_t$) and a standard treatment ($\mu_c$)
3. $k = n_c/n_t$: treatment allocation ratio
4. $\sigma$: variance (we assume variance is known when z-test and unknown t-test)
5. $\alpha$: type I error rate
6. $\beta$: type II error rate (Power: $1 - \beta$)

• Output:

1. $n_t$: sample size of treatment group
2. $n_c$: sample size of control group

13.2.3.3 An Example (Two-arms (Independent) Non-inferiority Test)

Suppose that the pharmaceutical company is interested in establishing non-inferiority of the test drug as compared to the active control agent. Similarly, we assume that the non-inferiority margin is chosen to be 5% ($\delta = 0.05$). Also, suppose the true difference in mean LDL between treatment groups is 0% ($\epsilon = \mu_t - \mu_c = 0$). Assuming that the standard deviation is 10% ($\sigma = 10\%$), with $\alpha = 0.05$, the sample sizes required for achieving an 80% power ($1 - \beta = 0.8$) can be obtained by the following steps,

• Select SAMPLE SIZE: Continuous Outcome.

• Select Number of Groups: Two (independent) and Test Objective: Non-inferiority.

• Input $\delta, \epsilon = \mu_t - \mu_c, \sigma, k, \alpha$ and $1 - \beta$.

• Click Submit.

Then the computed sample sizes in this situation are shown in Figure 13.7.
13.3 Two arms (paired)

13.3.1 Methods

Let $\epsilon_d = \mu_1 - \mu_2$ be the difference between the true mean response of two paired groups ($\mu_1$ and $\mu_2$). Without loss of generality, consider $\epsilon > 0$ ($\epsilon < 0$) an indication of improvement (worsening) of the test drug as compared to the reference value.

- **Hypothesis**: The hypothesis of interest is
  
  *(Two – sided)*)
  \[ H_0 : \epsilon_d = 0 \quad versus \quad H_1 : \epsilon_d \neq 0 \]

  *(One – sided)*)
  \[ H_0 : \epsilon_d \leq 0 \quad versus \quad H_1 : \epsilon_d > 0 \]
13.3. Two arms (paired)

13.3.3. An Example (Two-arms (paired) Equality Test)

- **Formula:** Denote $\Delta_d = \epsilon_d / \sigma_d$ be the effect size. And we use the **T-test** here to calculate $n$ that satisfies

  *(Two − sided)*

  $$T_{n-1} \{ t_{\alpha/2, n-1} | \sqrt{n} \Delta_d \} - T_{n-1} \{ -t_{\alpha/2, n-1} | \sqrt{n} \Delta_d \} = \beta$$

  *(One − sided)*

  $$T_{n-1} \{ t_{\alpha, n-1} | \sqrt{n} \Delta_d \} = \beta$$

13.3.2 Input and Output

- **Input:**
  - if we "Enter the effect size directly",
    1. $\Delta_d$: the effect size, could be calculated by $\Delta_d = (\mu_1 - \mu_2) / \sigma_d$, where $\mu_1$ and $\mu_2$ are mean response of two groups, and $\sigma_d$ is the standard deviation of pre-post difference
    2. $\alpha$: type I error rate
    3. $\beta$: type II error rate (Power: $1 - \beta$)
  - if "Calculate the effect size" is needed,
    1. $\mu_1$: mean response of group 1
    2. $\mu_2$: mean response of group 2
    3. $\sigma_d$: standard deviation of pre-post difference
    4. $\alpha$: type I error rate
    5. $\beta$: type II error rate (Power: $1 - \beta$)

- **Output:** $n$

13.3.3 An Example (Two-arms (paired) Equality Test)

Consider a standard two-period paired design for the trial whose objective is to establish therapeutic equality between a test drug and a standard therapy. The sponsor is interested in having an 80% ($1 - \beta = 0.80$) power for establishing therapeutic equality. Based on the results from previous studies, it is estimated that the variance is 20% ($\sigma_d = 0.20$). Suppose mean response of group 2 is 1.3 and mean response of group 1 is 1.2. That is, the true mean difference is 10% ($\mu_2(test) - \mu_1(reference) = 0.10$) and effect size $\Delta = 0.50$. The sample sizes can be obtained by the following steps,
Module 13. Sample Size Calculation for Continuous Outcome

- Select SAMPLE SIZE: Continuous Outcome.

- Select Number of Groups: Two (paired), 1 or 2 Sided Test: 2-Sided and Effect Size: Enter effect size directly.

- Input $\Delta_d, \alpha$ and $1 - \beta$.

- Click Submit.

or,

- Select Number of Groups: Two (paired), 1 or 2 Sided Test: 2-Sided and Effect Size: Calculate effect size $\Delta_d = |\mu_1 - \mu_2|/\sigma_d$.

- Input $\mu_1, \mu_2, \sigma_d, \alpha$ and $1 - \beta$.

- Click Submit.

Then the computed sample sizes in this situation are shown in Figure 13.8.

**Figure 13.8:** An Example (Two-arms (paired) Equality Test)
13.4 Multiple arms

13.4.1 Methods

Let \( x_{ij} \) be the \( j \)-th subject from the \( i \)-th treatment group, \( i = 1, \ldots, m \), \( j = 1, \ldots, n \). Consider the following one-way analysis of variance (ANOVA) model:

\[
x_{ij} = \mu_i + \epsilon_{ij},
\]

where \( \mu_i \) is the fixed effect of the \( i \)-th treatment and \( \epsilon_{ij} \) is a random error in observing \( x_{ij} \). It is assumed that \( \epsilon_{ij} \) are i.i.d. normal random variables with mean 0 and variance \( \sigma^2 \). Let

\[
SSE = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \mu_i)^2
\]

\[
SSA = \sum_{i=1}^{m} (\mu_i - \mu)^2,
\]

where

\[
\mu_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \quad \text{and} \quad \mu = \frac{1}{m} \sum_{i=1}^{m} \mu_i
\]

Then \( \sigma^2 \) can be estimate by

\[
\hat{\sigma}^2 = \frac{SSE}{m(n-1)}
\]

- **Hypothesis:** The hypothesis of interest is

\[
H_0 : \mu_1 = \mu_2 = \cdots = \mu_m \quad \text{versus} \quad H_1 : \mu_i \neq \mu_j \quad (1 \leq i \leq j \leq m).
\]

- **Formula:** Under the null hypothesis \( H_0 \), \( F_A = \frac{nSSA/(m-1)}{SSE/[m(n-1)]} \) follows F-distribution. So \( H_0 \) is rejected at the \( \alpha \) level of significance if

\[
F_A = \frac{nSSA/(m-1)}{SSE/[m(n-1)]} > F_{\alpha,m-1,m(n-1)}
\]

where \( F_{\alpha,m-1,m(n-1)} \) is the \( \alpha \) upper quantile of the F-distribution with \( m-1 \) and \( m(n-1) \) degrees of freedom.

Under the alternative hypothesis \( H_1 \), the power of this test is given by

\[
P(F_A > F_{\alpha,m-1,m(n-1)})
\]

Hence, the sample size needed to achieve power \( 1-\beta \) can be obtained by

\[
P(F_A > F_{\alpha,m-1,m(n-1)}) = 1 - \beta.
\]
13.4.2 Input and Output

- **Input:**
  1. $m$: number of groups
  2. $f$: effect size
  
  $$f = \frac{\sigma_m}{\sigma} = \sqrt{\frac{\sigma^2_m}{\sigma^2}}$$

  where $SSA/(m - 1)$ is approximately $\sigma^2_m$ and $SSE/m(n - 1)$ is approximately $\sigma^2$.
  3. $\alpha$: type I error rate
  4. $\beta$: type II error rate (Power: $1 - \beta$)

- If "Enter Effect Size Directly",

- If "Calculate Effect Size" is needed,
  1. $m$: number of groups
  2. $\mu_i$: mean of group $i$ ($1 \leq i \leq m$)
  3. $\sigma$: common standard deviation
  4. $\alpha$: type I error rate
  5. $\beta$: type II error rate (Power: $1 - \beta$)

- **Output:** $n$ for per group

13.4.3 An Example (Multiple-arms One-Way ANOVA Test)

Suppose that we are interested in conducting a four-arm ($m = 4$) parallel group, double-blind, randomized clinical trial to compare four treatments. The comparison will be made with a significance level of $\alpha = 0.05$. Assume that the standard deviation within each group is $\sigma = 3.5$ and that the true mean responses for the four treatment groups are given by,

$$\mu_1 = 8.25, \quad \mu_2 = 9.75, \quad \mu_3 = 9.00 \quad \text{and} \quad \mu_4 = 10.00.$$ 

Then, $f = 0.391$. The sample sizes required for achieving an 80% power ($1 - \beta = 0.8$) can be obtained by the following steps,

- Select **SAMPLE SIZE: Continuous Outcome**.
- Select **Number of Groups:** $> 2$ and **How to Determine Effect Size ($f$): Enter effect size directly**.
13.4. Multiple arms

13.4.3. An Example (Multiple-arms One-Way ANOVA Test)

- Input $m$, $f$, $\alpha$ and $1 - \beta$.
- Click Submit.

or,

- Select Number of Groups: $> 2$ and How to Determine Effect Size ($f$): Calculate effect size $f = \sigma_m / \sigma$.
- Input $m$, $\mu_i (i = 1, \ldots, 4)$, $\alpha$ and $1 - \beta$.
- Click Submit.

Then the computed sample sizes in this situation are shown in Figure 13.9.

**Figure 13.9:** An Example (Multiple-arms One-Way ANOVA Test)
14. Sample Size Calculation for Time-to-Event Outcome

In this section, we implement the sample size calculation for time-to-event endpoint, which include the following functions shown in Table 14.1.

<table>
<thead>
<tr>
<th>Number of Arms</th>
<th>Test Objectives</th>
<th>One- or/and Two-sided</th>
<th>Contents</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>Equality</td>
<td>One-sided &amp; Two-sided</td>
<td></td>
<td>Section 14.1</td>
</tr>
<tr>
<td>Two</td>
<td>Equality</td>
<td>One-sided &amp; Two-sided</td>
<td>Logrank test</td>
<td>Section 14.2</td>
</tr>
</tbody>
</table>

Before the text, there is three important symbols $A$, $F$ and $L$ for time-to-event endpoint introduced as follows:

![Diagram](image_url)

Accrual period $(A)$

Follow-up time $(F)$

Trial starts

Maximum follow-up time $(L = A + F)$

Trial ends
14.1 Single arm

In a study with a single arm, we assume for planning purposes that the survival times follow an exponential distribution with hazard \( h(t; \lambda) = \lambda \) and survival distribution \( S(t; \lambda) = e^{-\lambda t} \). After the trial is completed, we obtain a series of independent survival times \( t_1, t_2, \ldots, t_n \) and indicators \( \delta_1, \delta_2, \ldots, \delta_n \), with \( \delta_i = 1 \) for event occurring, \( \delta_i = 0 \) otherwise, where \( n \) is the total number of subjects in the trial. According to Moore (2016), \( \hat{\lambda} = d/V \), where

\[
d = \sum_{i=1}^{n} \delta_i \quad \text{and} \quad V = \sum_{i=1}^{n} t_i.
\]

14.1.1 Methods

- **Hypothesis**: The hypothesis of interest is

  (Two-sided)

  \[ H_0 : \lambda_t = \lambda_c \quad \text{versus} \quad H_1 : \lambda_t \neq \lambda_c, \]

  (One-sided)

  \[ H_0 : \lambda_t = \lambda_c \quad \text{versus} \quad H_1 : \lambda_t < \lambda_c, \]

  where \( \lambda_t \) and \( \lambda_c \) are the hazard rates for the current treatment and historical reference, respectively. The hypothesis is equivalent to

  (Two-sided)

  \[ H_0 : m_t = m_c \quad \text{versus} \quad H_1 : m_t \neq m_c, \]

  (One-sided)

  \[ H_0 : m_t = m_c \quad \text{versus} \quad H_1 : m_t > m_c, \]

  where \( m_t \) and \( m_c \) are median survival time for the current treatment and historical reference, respectively, or

  (Two-sided)

  \[ H_0 : HR = 1 \quad \text{versus} \quad H_1 : HR \neq 1, \]

  (One-sided)

  \[ H_0 : HR = 1 \quad \text{versus} \quad H_1 : HR < 1, \]

  where \( HR = \lambda_t/\lambda_c = m_c/m_t \) is the hazard ratio for the current treatment and historical reference.
• **Formula:** We can get sample size $n$ from

(Two − sided)

$$n_d = \frac{(z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

(One − sided)

$$n_d = \frac{(z_{\alpha} + z_{\beta})^2}{\Delta^2}$$

where $\Delta = \log(\lambda_t/\lambda_c)$ and $z_{\alpha}$ is the upper $\alpha$th quantile of the standard normal distribution, and

$$n = \frac{n_d}{P(\delta = 1)},$$

where $n_d$ is the number of event required, $n$ the total simple size required and the proportion of event occurring

$$P(\delta = 1) = 1 - \frac{1}{A\lambda_t} (e^{-\lambda_t F} - e^{-\lambda_t (A+F)}).$$

### 14.1.2 Input and Output

- **Input:**
  1. $m_c$ ($m_c = \frac{\log(2)}{\lambda_c}$): median survival time for historical control
  2. $m_t$ ($m_t = \frac{\log(2)}{\lambda_t}$): median survival time for treatment, or $HR (HR = m_c/m_t)$: hazard ratio
  3. $A$: length of accrual period
  4. $L$ ($L = A + F$): maximum follow-up time
  5. $\alpha$: type I error rate
  6. $\beta$: type II error rate (Power: $1 - \beta$)

- **Output:** number of event required $n_d$ and total simple size $n$

### 14.1.3 An Example (Single-arm One-sided Test)

Consider an example where we plan a single sample clinical trial with a 5% ($\alpha = 0.05$) significance level (one-sided) test, and we need 80% ($1 - \beta = 0.8$) power to detect a hazard ratio of 0.7 ($HR = 0.7$). Suppose that the null hypothesis rate is $m_c = 7$ months, and the alternative hypothesis hazard rate is $m_t = m_c/HR = 10$ months. We suppose now that the accrual period is $A = 3$ months and that the follow-up period is an additional $F = 6$ months (i.e., maximum follow-up time $L = 9$ months). To obtain an estimate of the number of patients, we follow these steps,
14.1. Single arm

14.2.1. Methods

- Select SAMPLE SIZE: Time To Event.

- Select Number of Groups: One, 1 or 2 Sided Test: 1-Sided, Time Unit: Months and Choose Input Mode: Hazard ratio and median survival time of historical control.

- Input $HR, m_c, A, L, \alpha$ and $1 - \beta$.

- Click Submit.

Then the computed sample size in this situation is shown in Figure 14.1.

**Figure 14.1:** An Example (Single-arm One-sided Test)
14.2 Two arms

14.2.1 Methods

- **Hypothesis:** The hypothesis of interest is
  
  *(Two − sided)*
  \[ H_0 : \lambda_t = \lambda_c \quad \text{versus} \quad H_1 : \lambda_t \neq \lambda_c, \]

  *(One − sided)*
  \[ H_0 : \lambda_t = \lambda_c \quad \text{versus} \quad H_1 : \lambda_t < \lambda_c, \]

  where \( \lambda_t \) and \( \lambda_c \) are the hazard rates for the current treatment and historical reference, respectively. The hypothesis is equivalent to

  *(Two − sided)*
  \[ H_0 : m_t = m_c \quad \text{versus} \quad H_1 : m_t \neq m_c, \]

  *(One − sided)*
  \[ H_0 : m_t = m_c \quad \text{versus} \quad H_1 : m_t > m_c, \]

  where \( m_t \) and \( m_c \) are median survival time for the current treatment and historical reference, respectively, or

  *(Two − sided)*
  \[ H_0 : HR = 1 \quad \text{versus} \quad H_1 : HR \neq 1, \]

  *(One − sided)*
  \[ H_0 : HR = 1 \quad \text{versus} \quad H_1 : HR < 1, \]

  where \( HR = \lambda_t/\lambda_c = m_c/m_t \) is the hazard ratio for the current treatment and historical reference.

- **Formula:** We can get sample sizes \( n_t \) and \( n_c \) from

  *(Two − sided)*
  \[ n_d = \frac{[(1 + k)(z_{\alpha/2} + z_\beta)]^2}{k\Delta^2} \]

  *(One − sided)*
  \[ n_d = \frac{[(1 + k)(z_\alpha + z_\beta)]^2}{k\Delta^2} \]

  where

1. \( \Delta = log(\lambda_t/\lambda_c) \).
14.2. Two arms

14.2.3. An Example (Two-arms One-sided Test)

2. \( k = n_t/n_c \)

3. \( z_\alpha \) is the upper \( \alpha \)th quantile of the standard normal distribution.

and

\[
n = \frac{n_d}{P(\delta = 1)}, \quad n_c = \frac{n}{1 + k} \quad \text{and} \quad n_t = \frac{kn}{1 + k},
\]

where \( n_d \) is the number of event required, \( n \) the total sample size required and \( P(\delta = 1) \) is the combined probability of event occurring. According to Schoenfeld (1983), we have

\[
P(\delta = 1) = \frac{P(\delta_c = 1)}{1 + k} + \frac{kP(\delta_t = 1)}{1 + k} = \frac{P(\delta_c = 1) + kP(\delta_t = 1)}{1 + k},
\]

where \( P(\delta_c = 1) \) and \( P(\delta_t = 1) \) are probabilities of event occurring for control and treatment, respectively, and are calculated as:

\[
P(\delta_c = 1) = 1 - \frac{1}{A\lambda_c} (e^{-\lambda_c F} - e^{-\lambda_c (A+F)}),
\]

\[
P(\delta_t = 1) = 1 - \frac{1}{A\lambda_t} (e^{-\lambda_t F} - e^{-\lambda_t (A+F)}).
\]

14.2.2 Input and Output

- **Input:**
  1. \( m_c \) (\( m_c = \frac{\log(2)}{\lambda_c} \)): median survival time for historical control
  2. \( m_t \) (\( m_t = \frac{\log(2)}{\lambda_t} \)): median survival time for treatment, or \( HR (HR = m_c/m_t) \): hazard ratio
  3. \( k \) (\( k = n_t/n_c \)): subject ratio of test control versus treatment
  4. \( A \): length of accrual period
  5. \( L \) (\( L = A + F \)): maximum follow-up time
  6. \( \alpha \): type I error rate
  7. \( \beta \): type II error rate (Power: \( 1 - \beta \))

- **Output:** number of event required \( n_d \), total sample size \( n \), sample size for control arm \( n_c \) and for test treatment \( n_t \)

14.2.3 An Example (Two-arms One-sided Test)

Consider a example where we plan a single sample clinical trial with a 5% (\( \alpha = 0.05 \)) significance level (one-sided) test, and we need 80% (\( 1 - \beta = 0.8 \)) power to detect a hazard ratio of 0.7 (\( HR = 0.7 \)).
0.7). Suppose that the null hypothesis rate is \( m_c = 7 \) months, and the alternative hypothesis hazard rate is \( m_t = m_c / HR = 10 \) months. We suppose now that the accrual period is \( A = 3 \) months and that the follow-up period is an additional \( F = 6 \) months (i.e., maximum follow-up time \( L = 9 \) months). To obtain an estimate of the number of patients with \( k = 1 \) (equal allocation), we follow these steps,

- Select **SAMPLE SIZE: Time To Event**.

- Select **Number of Groups: Two, 1 or 2 Sided Test: 1-Sided, Time Unit: Months** and **Choose Input Mode: Hazard ratio and median survival time of historical control**.

- Input \( HR, m_c, k, A, L, \alpha \) and \( 1 - \beta \).

- Click **Submit**.

Then the computed sample size in this situation is shown in Figure 14.2.
14.2. Two arms

14.2.3. An Example (Two-arms One-sided Test)

Figure 14.2: An Example (Two-arms One-sided Test)

Sample Size Calculation - Time To Event

<table>
<thead>
<tr>
<th>Number of Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>O One</td>
</tr>
<tr>
<td>O Two</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1 or 2 Sided Test</td>
</tr>
<tr>
<td>O 1-Sided</td>
</tr>
<tr>
<td>O 2-Sided</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Time Unit</td>
</tr>
<tr>
<td>O Months</td>
</tr>
<tr>
<td>O Years</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Choose input Mode</td>
</tr>
<tr>
<td>O Hazard ratio and median survival time of historical control</td>
</tr>
<tr>
<td>O Median survival time of historical control and treatment</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Hazard Ratio (HR=\frac{\hat{r}_2}{\hat{r}_1})</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>Median Survival Time for Historical Control (m2)</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>Subject Allocation Ratio (k = n1/n2)</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>Length of Accrual Period (A)</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>Maximum Follow-up Time (L)</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>Type I Error (α)</td>
</tr>
<tr>
<td>0.05</td>
</tr>
<tr>
<td>Power (1-β)</td>
</tr>
<tr>
<td>0.8</td>
</tr>
</tbody>
</table>

Result

Given an accrual period of 3 months, a maximum follow-up time of 9 months, in a one-sided test for two-sample time-to-event endpoint, at the significance level of 0.05, 195 events, 420 patients for total, 210 for control and 210 for treatment is needed to achieve 80% power when the hazard ratio is 0.7 and the median survival time for historical control is 7. And the proportion of event occurring is 0.463.
15. Simon’s Two-Stage Design

This section introduces the sample size calculation for Phase Ib/II clinical trial using Simon’s two-stage design (Simon, 1989).

15.1 Method

The Simon’s two-stage design is a one-sided one-sample design in which the treatment is tested against a historical control in its response rate. The hypothesis of interest in this design is

\[ H_0 : p \leq p_0 \quad versus \quad H_1 : p \geq p_1 \]

where \( p_0 \) is uninteresting response rate, which is often the historical response rate, and \( p_1 \) is expected response rate.

The design consists of two stages. In the first stage, \( n_1 \) patients will be recruited and treated and number of responses in the first stage \( (x_1) \) is assumed that \( x_1 \sim Bin(n_1, p) \). If there are \( r_1 \) or fewer responses among these \( n_1 \) patients, i.e., \( x_1 \leq r_1 \), the study will be early terminated and accept the null hypothesis. Otherwise, additional \( n_2 \) patients will be enrolled in the second stage and number of responses in the second stage \( (x_2) \) is assumed that \( x_2 \sim Bin(n_2, p) \), resulting in a total number sample size of \( n = n_1 + n_2 \). If there are less than or exactly \( r \) responses among these \( n \) patients, i.e., \( x = x_1 + x_2 \leq r \), we also accept the null hypothesis and claim that the treatment is not promising. The process of the design is shown in Figure 15.1.

2. Enumeration

For specified values of \( p_0, p_1 \), and type I/II error rates, \( \alpha \) and \( \beta \), we enumerate all of designs with

\[ n \in [1, n_{max}], \quad n_1 \in [1, n - 1], \quad r_1 \in [0, n_1] \quad and \quad r \in [r_1, n]. \]
We can get the expected sample size $EN = n_1 + (1 - PET)n_2$, where $PET$ represents the probability of early termination after the first stage and depends on the true probability of response $p$ (assumed as $p_0$):

$$PET = B(r_1; p_0, n_1) = \sum_{i=0}^{r_1} \binom{n_1}{i} p_0^i (1 - p_0)^{n_1-i},$$

where $B(*)$ denotes the cumulative binomial distribution. Then determine that

- **Optimal Two-stage Design**: satisfies the error probability constraints and minimizes the expected sample size ($EN$) when the response probability is $p_0$.

- **Minimax Two-stage Design**: satisfies the error probability constraints and minimizes the total sample size ($n$).

3. **Start of Enumeration**

The search over $n$ could be ranged from a lower value of about

$$\bar{p}(1 - \bar{p})\left[\frac{z_\alpha + z_\beta}{p_1 - p_0}\right]^2,$$

where $\bar{p} = (p_0 + p_1)/2$ and $z_\alpha$ is the upper $\alpha$th quantile of the standard normal distribution, to ensure that there are a nontrivial ($n_1, n_2 > 0$) two-stage design.

### 15.2 Program Input and Output

1. **Input**: $p_0$, $p_1$, $\alpha$, $\beta$, $n_{max}$.
   - $p_0$: uninteresting response rate or the historical response rate of the control
   - $p_1$: desirable target response rate
   - $\alpha$: type I error rate
   - $\beta$: type II error rate (Power: $1 - \beta$)
   - $n_{max}$: maximum sample size allowed when searching $n$

2. **Output**: $r_1, n_1, r, n, EN$ and $PET$ for the Optimal and Minimax designs.
   - $r_1$: the first stage threshold to stop the trial for futility, i.e., if there are $r_1$ or less responses, the trial will be early terminated.
   - $n_1$: the number of patients studied in the first stage.
Module 15. Simon’s Two-Stage Design

- $r$: 
- $n$: the total sample size.
- $PET$: the probability of early termination after the first stage under the null when the response probability is $p_0$.
- $EN$: the expected sample size, $EN = n_1 + (1 - PET)(n - n_1)$, under the null when the response probability is $p_0$.

15.3 Protocol Template

A Simons two-stage Optimal/(Minimax) design will be used to allow early stopping if the response is not sufficiently promising to warrant further development (i.e. $< p_0$). This design tests a null hypothesis that the true response rate is less than $p_0$ against a specific one-sided alternative hypothesis that the true response is at least $p_1$. The type I error rate is $\alpha$ (one-sided) and the type II error rate is $\beta$. Under these assumptions, a total of $n$ patients are planned for enrollment. Based on the above design considerations, $n_1$ patients will be enrolled to the first stage. If $\leq r_1$ patient in the cohort achieves a response, then enrollment will be early terminated. If at least $r_1 + 1$ patients achieve a response among the first $n_1$ patients, then an additional $n - n_1$ patients will be enrolled to the second stage. The null hypothesis will be rejected if at least $r + 1$ responses are observed among the $n$ patients.
Figure 15.1: Flow Chart of Simon’s Two Stage.
Reference


Lin, X. and Ji, Y. (2020b). The joint i3+3 (Ji3+3) design for phase I/II adoptive cell therapy clinical trials. *Contemporary Clinical Trials*, In press.


*Your trial designs anywhere, anytime.*


*Your trial designs anywhere, anytime.*


