BAYESIAN ADAPTIVE DESIGNS FOR EARLY PHASE CLINICAL TRIALS

Jiaying Guo

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Biostatistics,
Indiana University

July 2023

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____
Yong Zang, PhD, Chair

_____
Jiali Han, PhD

May 31, 2023

_____
Yi Zhao, PhD

_____
Jie Ren, PhD

DEDICATION

I dedicate my dissertation work to my parents, husband and daughter for their love and encouragement. I would not be where I am today, without the understanding and support from my family.

ACKNOWLEDGEMENT

First, I would like to thank my advisor, Dr. Yong Zang, for his instruction, guidance, advice, and great patience during my PhD study. His dedication, knowledge and insight in statistical and medical research is an inspiration to me. His mentorship and help have made this dissertation possible.

I also would like to give my sincere appreciation to my committee members, Dr. Jiali Han, Dr. Yi Zhao and Dr. Jie Ren for graciously serving on my committee and providing valuable advice on my research.

Finally, the faculty members, staff and fellow students in the Department of Biostatistics deserve my sincerest gratitude. I am truly humbled and honored to work with you all.

Thank you all for being there with me all the way.

Jiaying Guo

BAYESIAN ADAPTIVE DESIGNS FOR EARLY PHASE CLINICAL TRIALS

Delayed toxicity outcomes are common in phase I clinical trials, especially in oncology studies. It causes logistic difficulty, wastes resources, and prolongs the trial duration. We propose the time-to-event 3+3 (T-3+3) design to solve the delayed outcome issue for the 3+3 design. We convert the dose decision rules of the 3+3 design into a series of events. A transparent yet efficient Bayesian probability model is applied to calculate the event happening probabilities in the presence of delayed outcomes, which incorporates the informative pending patients' remaining follow-up time into consideration. The T-3+3 design only models the information for the pending patients and seamlessly reduces to the conventional 3+3 design in the absence of delayed outcomes. We further extend the proposed method to interval 3+3 (i3+3) design, an algorithm-based phase I dose-finding design which is based on simple but more comprehensive rules that account for the variabilities in the observed data. Similarly, the dose escalation/de-escalation decision is recommended by comparing the event happening probabilities which are calculated by considering the ratio between the averaged follow-up time for at-risk patients and the total assessment window. We evaluate the operating characteristics of the proposed designs through simulation studies and compare them to existing methods.

The umbrella trial is a clinical trial strategy that accommodates the paradigm shift towards personalized medicine, which evaluates multiple investigational drugs in

different subgroups of patients with the same disease. A Bayesian adaptive umbrella trial design is proposed to select effective targeted agents for different biomarker-based subgroups of patients. To facilitate treatment evaluation, the design uses a mixture regression model that jointly models short-term and long-term response outcomes. In addition, a data-driven latent class model is employed to adaptively combine subgroups into induced latent classes based on overall data heterogeneities, which improves the statistical power of the umbrella trial. To enhance individual ethics, the design includes a response-adaptive randomization scheme with early stopping rules for futility and superiority. Bayesian posterior probabilities are used to make these decisions. Simulation studies demonstrate that the proposed design outperforms two conventional designs across a range of practical treatment-outcome scenarios.

Yong Zang, PhD, Chair

Jiali Han, PhD

Yi Zhao, PhD

Jie Ren, PhD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DLT - Dose-Limiting Toxicity

MTD - Maximum Tolerated Dose

FDA - Food and Drug Administration

EMA - European Medicines Agency

irAE - immune-related Adverse Event

BOIN - Bayesian Optimal Interval Design

AFR - Averaged Follow-up Ratio

PAVA - Pooled-Adjacent-Violators Algorithm

PCS - Percentage of Correct Selection

CBR - Clinical Benefit Rate

PFS - Progression-Free Survival

IUSCCC - Indiana University Simon Comprehensive Cancer Center

NSCLC - Non-Small-Cell Lung Cancer

EGFR - Epidermal Growth Factor Receptor

KRAS - Kirsten Rat Sarcoma Virus

RECIST - Response Evaluation Criteria In Solid Tumors

BDA - Bayesian Data Augmentation

CR - Complete Response to Treatment

PR - Partial Response to Treatment

SD - Stable Disease

ULCM - Umbrella design using the Latent Class Mode

Chapter 1

Introduction

Most phase I dose finding clinical trials rely on ascertaining the toxicity outcomes of currently treated patients prior to determining dosing assignment in upcoming patients. However, delayed toxicity is common in oncology clinical trials and other studies, which happens when there are patients in the trial with pending dose-limiting toxicity (DLT) while the next patient or group is ready for dose assignment. It wastes resources and causes a tremendous administrative burden. Various dose finding trials designed to accommodate delayed outcomes have been proposed. However, these designs require sophisticated statistical approaches and were only applicable for model-based and model-assistant designs, limiting its clinical application. The 3+3 design is arguably the most popular phase I clinical trial design due to its simplicity, transparency, and costless implementation (Storer, 1989; Storer, 2001). However, delayed outcome issue for the 3+3 design has not been adequately addressed in a statistically rigorous manner. In Chapter 2, we propose a new phase I trial design as the extension of the 3+3 design handling the delayed outcome issue. It is derived based on the decision rules of the 3+3 design and incorporate the information from the pending patients' remaining follow-up time through a statistically rigorous Bayesian probability model.

Although 3+3 design is the most commonly used design in phase I clinical trial, it is widely criticized for its unsafe and unreliable operating characteristics. Another model-assisted design i3+3 (interval 3+3) which is based on simple but more comprehensive

rules that account for the variabilities in the observed data shows superior to the 3+3 design in trial safety and the ability to identify the true MTD (Liu *et al*., 2020). To deal with delayed toxicity issue in i3+3 design, in Chapter 3, we propose a new design based on i3+3, in which a Bayesian probability model is applied to calculate the event happening probabilities in the presence of delayed outcomes. The dose escalation/de-escalation decision is made by comparing the event happening probabilities which account for the ratio between the averaged follow-up time for at-risk patients and the total assessment window.

Recent advances in biomedicine and genomics have brought better understanding of cancer-causing mechanisms and the ability to identify the corresponding therapeutic targets. However due to the heterogeneous biological profiles of patients, specific targeted agents may not be effective for the general population of patients, but work for only a small proportion of patients, and some agents may not work well at all (Mandrekar and Sargent, 2009; Freidlin et al., 2010; Sargent and Mandrekar, 2013). Unlike the traditional one-drug-one-indication approach, the umbrella trial evaluates multiple investigational drugs in patients with the same cancer but different gene mutations (changes) or biomarkers. However, challenges exist when designing these phase II oncology trials. Defining progression-free survival (PFS) as the primary outcome often requires longer follow-up time. This is the first challenge because extended follow-up time can impede the implementation of interim analysis. The second challenge is the difficulty of obtaining adequate sample size due to the slow accrual rates. As a result, detecting subgroup-specific treatment effects becomes unlikely when subgroups are

separately analyzed. To address these challenges, in Chapter 4, we propose a Bayesian adaptive umbrella trial design by using a surrogate marker for the long-term PFS, the clinical benefit rate (CBR), which can be assessed much sooner than PFS. We propose a mixture regression model with the purpose of borrowing information between the CBR and PFS. A data-driven Bayesian latent class model is then used to adaptively group individual subgroups into latent classes based on their overall data heterogeneities. The subgroup-specific sample size limitation is thus resolved by pooling response outcomes within the latent classes.

In Chapter 5, we discuss the strengths and limitations of our proposed designs and possible extensions of our work.

Chapter 2

T-3+3: 3+3 Design with Delayed Outcomes

2.1 Introduction

The primary goal of a phase I dose-finding clinical trial is to identify the maximum tolerated dose (MTD), which is defined as the dose with a toxicity rate closest to a prespecified target. These dose-finding trials are critically important as they determine at which dose level the drug will be further investigated in subsequent phase II and phase III trials.

The majority of phase I clinical trials require that toxicity outcomes can be observed quickly such that all the toxicity outcomes of the currently treated patients are observed by the time of dose assignment for the coming patients. However, delayed toxicity outcomes, which violate this requirement, are common in clinical trials, especially in oncology studies. For example, in radiotherapy trials, dose-limiting toxicities (DLTs) often occur long after the treatment is finished (Muler *et al.*, 2004; Desai *et al.*, 2007). In immunotherapy trials, Ghisoni et al. (2021) recently reviewed all registration trials leading to Immune checkpoint inhibitors (ICI)'s approval by the US FDA and EMA up to December 2019. In their real-world analysis, for a total of 318 reported immune-related adverse events (irAEs), about 40% were ongoing at a median follow-up of 369 days (95% CI [194, 695]) or patient death.

The naive method ignoring the delayed toxicity outcomes may underestimate the toxicity probabilities, which would cause an undesirably large number of patients to be

treated at overly toxic doses and is not practical. A “safer” approach would be to suspend the accrual and wait for each patient's toxicity outcome to be fully observed prior to the next dose assignment, which may result in an unusually lengthy trial duration. Moreover, frequently suspending patient accrual waste resources also causes a tremendous administrative burden.

Many phase I dose-finding trial designs handling delayed outcomes have been proposed. Cheung and Chappell (2000) proposed a weighted likelihood method that weighted the likelihood of each patient by his/her actual follow-up time. Lin and Yuan (2020) proposed a binomial approximation approach to facilitate the computation of the weighted likelihood. Yuan and Yin (2011) developed an EM algorithm to impute the missing outcomes from the observed information, and this idea was further extended by Liu et al. (2014) to develop a data augmentation method under the Bayesian framework. Yuan et al. (2018) also proposed a single imputation approach to address the delayed outcome problem for the BOIN design. However, all these designs were developed using sophisticated statistical approaches and were only applicable for model-based and model-assistant designs (e.g., CRM, BOIN, Keyboard, etc.) (O’Quigley *et al*., 1990; Liu and Yuan, 2015; Yan *et al*., 2017).

The 3+3 design is arguably the most popular phase I clinical trial design (Storer, 1989; Storer, 2001). The reasons for the popularity of the 3+3 design include simplicity, transparency, and costless implementation. Indeed, the 3+3 design has been used for over 95% of published phase I oncology trials in the past two decades (Rogatko *et al.*, 2007; Tourneau *et al*., 2009), indicating a leading position in dose-finding oncology trial

designs. However, no statistically rigorous method is being developed to accommodate the delayed outcome issue for the 3+3 design. One possible explanation is that all the existing statistical methods handling delayed outcomes require an advanced statistical modeling approach. In contrast, the 3+3 design is a model-free algorithm-based design with the dose escalation/de-escalation rules determined by a prespecified empirical algorithm. To the best of our knowledge, the rolling six design is the only existing algorithm-based phase I clinical trial design handling delayed outcome (Skolnik *et al.*, 2008).

The rolling six design allows for accrual of two to six patients concurrently onto a dose level based on the number of patients currently enrolled, the number of patients experiencing DLT, and the number of patients still at risk, and therefore avoid the trial suspension caused by the delayed outcome. However, like most algorithm-based designs, the decision rule of the rolling six design is entirely empirical without the support of a rigorous statistical theorem. Therefore, the rolling six design cannot consider the remaining follow-up time for at-risk patients, which is critically informative and strongly correlated with the chance of developing a DLT. For example, assuming two patients, A and B, both are under follow-up in a trial and are at risk of developing a DLT, with an assessment window of three months. Patient A has been followed for just one week, and patient B has been followed for 11 weeks. Then, it is evident that compared with patient A, the probability of developing DLT for B is much lower because this patient is almost at the end of the assessment window. However, the rolling six design ignores this valuable information and treats patients A and B identically in its decision rules. Therefore, as

shown in our computer simulation studies, the performance of the rolling six design is barely satisfactory.

We propose a new phase I trial design as the extension of the 3+3 design handling the delayed outcome issue, referred to as the T-3+3 design. It is derived based on the decision rules of the 3+3 design and therefore enjoys the same appealing properties such as simplicity and transparency. Like the rolling six design, the T-3+3 design avoids unnecessary trial suspension and significantly shortens the trial duration. Also, like all the algorithm-based designs, the T-3+3 is costless implementable because its decision rule can be pre-tabulated, which dramatically facilitates the trial preparation and conduct without intensive computation and statistical consultation. Nevertheless, unlike the rolling six design, the T-3+3 is derived based on a statistically rigorous Bayesian probability model, enabling it to incorporate the information from the pending patients' remaining follow-up time. As a result, the T-3+3 design is more efficient in MTD identification and patient allocation.

2.2 Methods

2.2.1 Overview of the Delayed Outcome and the 3+3 Design

We first describe the delayed outcome issue for phase I clinical trials. In the phase I clinical trial, each patient is followed for a fixed period (e.g., assessment window) to score the DLT. Hence, if the patient has experienced the DLT at any time of the assessment window, the DLT score will be 1; otherwise, the DLT score will be 0. An interim dose decision will be made for new patients when a pre-determined number of patients have

been treated at the current dose level (3 or 6 for the 3+3 design). Therefore, if the patients' accrual rate cannot catch up with the assessment window, some patients in the trial may still be under follow-up and are at risk at the interim analysis time, which may or may not experience DLTs in the remaining follow-up time, so their DLT scores are missing. The delayed outcome issue arises because the DLT information for all the previous patients is required to determine the next dose.

We next briefly review the 3+3 design. There are many versions of 3+3 designs in the literature (Ji and Wang, 2013), and we adopt the one used by the rolling six design. We note that the T-3+3 design is general and can be equipped with any version of the 3+3 design along the same line. For a 3+3 design, when a total of 3 patients have been treated at the current dose, the next cohort of 3 patients will be treated at one dose higher, the current dose or one dose lower if the number of patients experiencing DLT is 0, 1, or at least 2 respectively. When a total of 6 patients have been treated at the current dose, the next cohort of 3 patients will be treated at one dose higher (if that dose has not been tried) or one dose lower if the number of patients experiencing DLT is at most 1 or at least 2 respectively. Up to 6 patients can be treated at one dose. At the end of the trial, The MTD is defined as the dose at which at most 1 of 6 participants experience a DLT, when at least 2 of 3 to 6 participants experience a DLT at one dose higher.

In summary, in the 3+3 design, the dose decision is based on the number of patients experiencing DLTs at the current dose level, referred to as r. If a total of 3 patients have been treated, then we have the dose-escalation event $E_3 \equiv \{r = 0\}$, dose-retaining event $R_3 \equiv \{r = 1\}$ and dose-deescalation event $D_3 \equiv \{r \geq 2\}$. Similarly, if a total of 6

patients have been treated, then we have the dose-escalation event $E_6 \equiv \{r = 1\}$ and the dose-deescalation event $D_6 \equiv \{r \geq 2\}$. Without the delayed outcomes, it is immediately clear which event is happening, so the corresponding dose decision can be made. However, the delayed outcome adds the uncertainty of the event identification, which subsequently affects the dose decision. For example, if at a dose assignment time, among 3 patients treated in the current dose, 1 has experienced DLT, 1 has completed follow-up, and does not experience DLT. The last one is at risk of experiencing DLT but is still under follow-up. Then, we may observe either R_3 or D_3, so the possible actions are to (1) retain at the current dose if we believe R_3 will eventually happen, (2) de-escalate if we believe D_3 will eventually happen and (3) suspend the trial to wait for more information. By taking action (1) or (2), we do not need to suspend the trial but are at the risk of making the wrong decision violating the 3+3 design. We can always take the correct action by taking action (3), but the trial duration will be prolonged. This is the dilemma we need to solve.

2.2.2 Probability Model and Dose-finding Algorithm

Our idea is that although we cannot completely avoid the event identification error in the presence of delayed outcomes, we still can derive the event happening probability by utilizing the DLT information and the remaining follow-up time from the pending patients. Then, the most likely happening event can be picked through the maximum probability for all the possible events. After that, we just need to compare that

probability with a prespecified cut-off value to determine whether we have enough information to continue accrual without trial suspension.

Specifically, we define that a total of n patients have been treated at the current dose. Among them, n_1 patients have completed the follow-up for DLT assessment, and r_1 of them have experienced DLT. The other $n_2 = n - n_1$ patients are at risk of delayed outcomes. We further define U as the fixed assessment window (e.g., 3 months) and v_i as the actual follow-up time (e.g., 1 week) for the ith at-risk patient for $i = 1, \cdots, n_2$ such that $U - v_i$ is the remaining follow-up time. Then, after denoting $\overline{v} = \frac{\sum_{i=1}^{n_2} v_i}{n_2}$ as the averaged follow-up time for all the patients at risk at the current dose, the distribution for the number of patients who will eventually experience DLT among the n2 pending patients, referred to as r_2, can be expressed as the following beta-binomial distribution (see Appendix for detailed derivation):

$$r_2 \sim Beta - binomial(n_2, k, r_1 + 1, n_1 - r_1 + n_2 \frac{\overline{v}}{U} + 1) \quad (1)$$

Therefore, let $\mathcal{D} = \{r_1, n_1, \overline{v}, n_2\}$ be the observed data, the probability of getting $r_2 = k$ given $\mathcal{D}$, referred as $\pi(k|\mathcal{D})$, can be written as

$$\pi(k|\mathcal{D}) = \frac{n_2!}{k!(n_2-k)!} \frac{\mathrm{B}(k+r_1+1, n_2-k+n_1-r_1+n_2\frac{\overline{v}}{U}+1)}{\mathrm{B}(r_1+1, n_1-r_1+n_2\frac{\overline{v}}{U}+1)} \quad (2)$$

with $\mathrm{B}(\cdot)$ denoting the beta function. Finally, by using formula (2), we can easily calculate the probability for each event happening, referred to as $\pi(D_3|\mathcal{D})$, $\pi(R_3|\mathcal{D})$, $\pi(E_3|\mathcal{D})$, $\pi(D_6|\mathcal{D})$ and $\pi(E_6|\mathcal{D})$.

We further define ϕ_D, ϕ_R and ϕ_E as 3 cut-off values for necessary trial suspension due to the lack of information. That is, if a dose de-escalation is recommended by

comparing the event happening probabilities, we will need to further compare $\pi(D_3|\mathcal{D})$ (or $\pi(D_6|\mathcal{D})$) with the cut-off value ϕ_D. We will de-escalate the dose if $\pi(D_3) > \phi_D$. However, suppose $\pi(D_3) \leq \phi_D$ indicates that we do not have enough information to support the action of dose de-escalation. In that case, we will need to suspend the trial to accumulate more data. Similar procedures will be done for dose escalation and dose retaining, with ϕ_E and ϕ_R being used separately. In addition, to further strengthen the trial safety, we require trial suspension if the latest cohort of 3 patients are all at risk with pending outcomes during a dose assignment time.

A key innovation of the T-3+3 design is that its dose escalation/de-escalation rule can be pre-tabulated and included in the trial protocol before the trial starts, which dramatically simplifies design implementation. The key is to link the averaged follow-up ratio (AFR), $\frac{\overline{v}}{U}$, which is the ratio between the averaged follow-up time for at-risk patients at current dose $\overline{v}$ and the total assessment window U, with the probabilities of events happening. When the AFR is given, together with r_1, n_1, and n_2, which are all immediately observable, we can calculate each event happening probability and determine the dosing movement. T-3+3 design also enjoys the desirable feature that its decision rule is invariant to the length of the assessment window because AFR is a measurement standardized by U. This means that the same decision table can be used to guide dose escalation and de-escalation, regardless of the length of the assessment window.

The cut-off values ϕ_D, ϕ_R and ϕ_E need further discussion. Small cut-off values will decrease the chance of trial suspension and shorten the trial duration. However, as a trade-off, less information will be collected, so the probability of making inappropriate

dose decisions (e.g., escalating the dose with 2 underlying DLTs) will increase. On the other side, large cut-off values will make the T-3+3 design close to the 3+3 design and impede the benefits of time/resource saving. We propose solving this dilemma through computer simulation studies and selecting the best values balancing the MTD selection, patient allocation, and trial duration. Based on our preliminary simulation studies, we recommend using $\phi_E = 0.5$, $\phi_R = 0.5$ and $\phi_D = 0.75$ as the default values.

Table 2.1 shows the dose decision rule for the T-3+3 design with the default cut-off values. At each dose assignment time, at the current dose, we need to count the number of patients, the number of DLT, the number of at-risk patients with pending outcomes, and their AFR. Then, we can easily use Table 2.1 to make the dose decision for the future cohort of patients without unnecessary trial suspension. Suppose that 3 patients have been treated at the current dose, 2 have completed 90 days follow-up, and 1 is still pending, and no DLT has been identified. We use decision rule 1 to escalate the dose regardless of the AFR. Consider another example that still 3 patients have been treated, 1 has experienced DLT, and the other two pending patients have been followed for 15 and 30 days, respectively. We need to apply decision rule 4 and calculate $\mathrm{AFR} = (15 + 30)/(2 * 90) = 0.25$. Then, since $\mathrm{AFR} < 0.295$, we should de-escalate the dose.

Table 2.1 Dose decision rule for T-3+3 design with $\phi_E = 0.5$, $\phi_R = 0.5$ and $\phi_D = 0.75$

Decision rule	No. of patients	No. of DLTs	No. of pending	Action			
				Escalate	Retain	De-escalate	Suspend
1	3	0	≥2	Yes			
2	3	0	3				Yes
3	3	1	≤1		Yes		
4	3	1	2			AFR ≤0.295	AFR >0.295
5	3	≥2	Any			Yes	
6	6	0	Any	Yes			
7	6	1	≤1	Yes			
8	6	1	2	AFR >0.187			AFR ≤0.187
9	6	1	3				Yes
10	6	1	4			AFR ≤0.386	AFR >0.386
11	6	1	5			AFR ≤0.707	AFR >0.707
12	6	≥2	Any			Yes	

As an extension of the 3+3 design, the T-3+3 design also follows the general requirement that no more than 6 patients can be treated at one dose level. Therefore, if on the dose selected for the next cohort of patients, there are already 6 patients being treated, the trial should be stopped. Also, additional modification is needed for doses on the boundary. If the lowest dose is the current dose and the decision is to de-escalate, all the doses under investigation are overly toxic. Then, we should terminate the trial and claim that no dose should be selected for further study. If the highest dose is the current dose and the decision is to escalate, we should treat 3 more patients on the highest dose. Then, if 2 or more of the 6 patients experience a DLT, we should de-escalate the dose. Otherwise, we stop the trial.

At the end of the trial, we finish the follow-up for each patient in the trial and get the complete DLT data. Then, a final step is needed to select the MTD. For the 3+3 design without the delayed outcome issue, the estimates of the DLT rates are monotonically ordered. That is, let $\{\hat{p}_j\}$ be the DLT rates estimates for all the doses under investigation $j = 1, 2, \ldots, J$. Then, we automatically have $\hat{p}_1 \leq \hat{p}_2 \leq \ldots \leq \hat{p}_J$ if a 3+3 design is used and there is no delayed outcomes happening. Consequently, the MTD is immediately clear. However, delayed outcomes add additional variability to the T-3+3 design because the dose escalation/de-escalation rule is based on the event probability rather than the complete observation of the event. Therefore, the monotonic order of $\{\hat{p}_j\}$ does not hold for the T-3+3 design. To solve this problem, we propose to use the isotonic estimates $\{\tilde{p}_j\}$, which can be obtained by applying the fast and efficient pooled-adjacent-violators algorithm (PAVA) (Brink *et al.*, 1973) to $\{\hat{p}_j\}$. In particular, the PAVA replaces any adjacent

$\hat{p}_j$'s that violate the non-decreasing order by their (weighted) average so that the resulting estimates $\tilde{p}_j$'s become monotonic. Then, the dose with the isotonic estimate $\tilde{p}_j$ closest to the targeted DLT rate ϕ is selected as the MTD. Due to the isotonic regression method we use for final MTD selection, the T-3+3 design is more flexible than the 3+3 and rolling six designs because it can target any prespecified DLT rate. We will provide freely available software associated with this paper to calculate the isotonic estimates and select the MTD.

2.3 Computer simulations

2.3.1 Simulation Setup

We used computer simulations to compare the 3+3, rolling six, and T-3+3 designs. We considered a phase I clinical trial with 6 doses. The DLT assessment window was 3 months. The patient arrival time was simulated from an exponential distribution with the accrual rate 2 patients per month. The time-to-DLT outcomes were simulated from Weibull distribution by controlling that 50% of the DLT occurred in the latter half of the assessment window. We considered the targeted DLT rate of 0.3, which was widely used for 3+3 design. As mentioned earlier, 3+3 and rolling six designs cannot target any prespecified DLT rate, but the T-3+3 design needs this value for final MTD selection. We considered 5 performance metrics based on 10,000 simulated trials.

1. Percentage of correct selection (PCS) of the MTD.
2. Percentage of patients allocated to the MTD.
3. Average trial duration.

4. Average number of patients.
5. Average number of DLTs.

Metrics 1 and 2 measure the accuracy of the design; metrics 3 and 4 measure the cost-effectiveness; and metric 5 measures the safety.

2.3.2 Simulation Results

Table 2.2 summarized the computer simulation results under 8 representative scenarios with various dose-toxicity curves. The MTD selection percentage under the dose level "0" indicates the percentage of early terminating the trial for overly toxic. The numbers in bold were the PCS for different designs. Based on Table 2.2, the T-3+3 design generally outperformed the 3+3 design and rolling six design in terms of metrics 1 (PCS) and 3 (duration), and the improvements were substantial. For example, in scenario 1, dose 6 is the true MTD with a DLT rate of 0.32. The T-3+3 design yielded a PCS of 35.6%, which was 13.7% greater than the 3+3 design and 16.8% greater than the rolling six design. Meanwhile, the T-3+3 design also yielded the shortest trial duration of 20.7 months, which was 9.7 months less than the 3+3 design and 5.9 months less than the rolling six design. Consider scenario 7 as another example which contained two MTD dose 3 and dose 4 (equal difference to targeted value 0.3). The PCS for T-3+3, 3+3, and rolling six designs were 45.3% VS 32.3% VS 29.5%, and the counterpart of trial duration was 14.4 months VS 21.1 months VS 15.6 months. For metrics 2, 4, and 5, the T-3+3 and 3+3 designs yielded similar performances and, in general, outperformed the rolling six design. Take scenario 1 as an example again. The percentages of patients allocated to the MTD (dose

6) were 13.5% under the T-3+3 design and 13.7% under the 3+3 design. The number decreased to 9.6% under the rolling six design. The average number of patients was 19.9 under the T-3+3 design and 21.5 under the 3+3 design and significantly increased to 29.5 under the rolling six design. Finally, all the designs reported a similar average number of DLTs ranging from 2.5 to 3.4. Nevertheless, the T-3+3 design took the minimum, and the rolling six design took the maximum.

We conducted additional simulations to examine the robustness of the T-3+3 design in terms of accrual rate, percentage of toxicity occurring in the latter half of the assessment window, and time-to-toxicity distribution. Due to page limits, we focused on metrics 1 and 2 for this sensitivity analysis. Figures 2.1 to 2.3 summarized the results for metric 1 (PCS), and Figures 6.1 to 6.3 summarized the results for metric 2 (percentage of patient allocation). The results were consistent with those reported above, which confirmed that the T-3+3 design was robust and yielded desirable operating characteristics under various settings.

Table 2.2: The operating characteristics of the 3+3, Rolling 6 and T-3+3 designs, including percentages of MTD selection (Selection (%)) and patients' allocation (Patients (%)) at each dose level, number of patients in the trial and trial duration, based on 10,000 simulated trials.

Designs		Dose level							Duration (month)	No. of Patients	No. of DLT
		0	1	2	3	4	5	6			
					Scenario 1						
	True Toxicity		0.05	0.06	0.08	0.11	0.19	**0.32**			
3 + 3	Selection(%)	2.80	3.5	5.9	10.5	23.5	32.0	**21.9**	30.4	21.5	2.8
	Patients (%)		16.3	16.4	17.0	18.2	18.4	13.7			
Rolling 6	Selection(%)	3.1	4.7	6.8	11.3	24.6	30.9	**18.8**	26.6	29.5	3.4
	Patients (%)		20.3	19.6	18.6	17.2	14.7	9.6			
T-3+3	Selection(%)	2.1	3.1	6.3	8.8	15.7	28.3	**35.6**	20.7	19.9	2.5
	Patients (%)		17.3	17.2	17.1	17.5	17.4	13.5			
					Scenario 2						
	True Toxicity		0.06	0.08	0.12	0.18	**0.30**	0.41			
3 + 3	Selection(%)	3.7	6.5	11.5	21.3	31.1	**19.3**	6.6	27.6	19.8	3.2
	Patients (%)		18.5	19.0	20.0	19.9	15.2	7.3			
Rolling 6	Selection(%)	4.6	7.1	14.1	21.8	30.6	**17.4**	4.4	22.6	25.5	3.6
	Patients (%)		23.4	22.3	20.6	17.1	11.8	4.7			
T-3+3	Selection(%)	3.1	4.4	9.9	17.0	27.0	**25.9**	12.8	18.6	18.3	2.9
	Patients (%)		19.6	19.7	20.1	19.2	14.4	7.1			
					Scenario 3						
	True Toxicity		0.05	0.10	0.20	**0.31**	0.50	0.70			
3 + 3	Selection(%)	2.6	9.7	27.5	34.3	**22.6**	3.4	0.1	23.6	17.3	3.3
	Patients (%)		21.2	24.9	25.9	18.9	7.9	1.2			
Rolling 6	Selection(%)	3.3	11.1	28.7	35.0	**19.5**	2.5	0.0	18.6	21.3	3.6
	Patients (%)		28.0	27.0	23.5	15.2	5.6	0.6			
T-3+3	Selection(%)	2.1	5.2	18.4	33.3	**31.3**	9.2	0.5	16.1	16.2	3.1
	Patients (%)		22.3	25.2	24.9	18.2	8.1	1.3			
					Scenario 4						
	True Toxicity		0.08	0.15	**0.29**	0.43	0.50	0.57			
3 + 3	Selection(%)	6.7	19.2	38.6	**27.5**	6.8	1.1	0.1	20.3	15.1	3.2
	Patients (%)		27.1	30.8	26.1	12.4	3.2	0.5			
Rolling 6	Selection(%)	8.0	20.1	40.2	**25.4**	5.6	0.7	0.0	15.2	17.7	3.4
	Patients (%)		33.7	30.8	23.4	9.9	1.9	0.2			
T-3+3	Selection(%)	5.7	9.4	31.0	**36.8**	13.8	2.8	0.5	13.9	14.2	3.1
	Patients (%)		28.0	30.2	25.1	12.5	3.5	0.7			

Table 2.2 continues.

Designs		Dose level 0	1	2	3	4	5	6	Duration (month)	No. of Patients	No. of DLT
					Scenario 5						
	True Toxicity		0.13	**0.28**	0.41	0.50	0.60	0.70			
3 + 3	Selection(%)	16.0	42.5	**30.8**	9.0	1.6	0.0	0.0	16.1	12.3	3.2
	Patients (%)		39.8	36.1	18.1	5.1	0.8	0.0			
Rolling 6	Selection(%)	18.1	44.5	**29.3**	7.2	0.9	0.0	0.0	10.9	13.2	3.2
	Patients (%)		44.9	35.7	15.7	3.3	0.4	0.0			
T-3+3	Selection(%)	13.8	22.4	**42.2**	17.2	4.1	0.4	0.0	11.1	11.7	3.0
	Patients (%)		40.0	35.1	18.1	5.6	1.1	0.1			
					Scenario 6						
	True Toxicity		**0.32**	0.49	0.56	0.65	0.76	0.87			
3 + 3	Selection(%)	61.0	**33.1**	5.4	0.5	0.0	0.0	0.0	9.3	7.6	2.9
	Patients (%)		67.7	27.1	4.7	0.5	0.0	0.0			
Rolling 6	Selection(%)	61.8	**33.7**	4.2	0.3	0.0	0.0	0.0	6.0	8.0	3.0
	Patients (%)		71.0	25.8	3.0	0.2	0.0	0.0			
T-3+3	Selection(%)	55.6	**28.9**	13.7	1.7	0.1	0.0	0.0	6.8	7.5	2.9
	Patients (%)		65.4	27.8	5.9	0.9	0.1	0.0			
					Scenario 7						
	True Toxicity		0.08	0.16	**0.28**	**0.32**	0.40	0.52			
3 + 3	Selection(%)	6.3	21.0	36.0	**20.6**	**11.7**	3.8	0.5	21.1	15.6	3.2
	Patients (%)		26.4	29.7	23.8	12.7	5.7	1.6			
Rolling 6	Selection(%)	8.2	22.6	36.8	**19.5**	**10.0**	2.7	0.3	15.6	18.1	3.4
	Patients (%)		32.9	30.0	22.1	10.2	4.0	0.9			
T-3+3	Selection(%)	5.1	10.1	30.6	**28.7**	**16.6**	7.3	1.6	14.4	14.7	3.0
	Patients (%)		27.2	29.1	22.9	12.8	6.0	1.9			
					Scenario 8						
	True Toxicity		0.44	0.58	0.63	0.79	0.84	0.87			
3 + 3	Selection(%)	**82.2**	16.5	1.2	0.1	0.0	0.0	0.0	6.8	5.9	2.7
	Patients (%)		80.8	17.5	1.6	0.1	0.0	0.0			
Rolling 6	Selection(%)	**82.7**	16.5	0.8	0.0	0.0	0.0	0.0	4.4	6.4	3.0
	Patients (%)		85.3	14.0	0.7	0.0	0.0	0.0			
T-3+3	Selection(%)	**76.4**	18.8	4.6	0.2	0.0	0.0	0.0	5.3	6.0	2.8
	Patients (%)		77.4	20.1	2.4	0.1	0.0	0.0			

Figure 2.1: The correct MTD selection percentages of the 3+3, Rolling 6, and T-3+3 designs with different patient accrual rates.

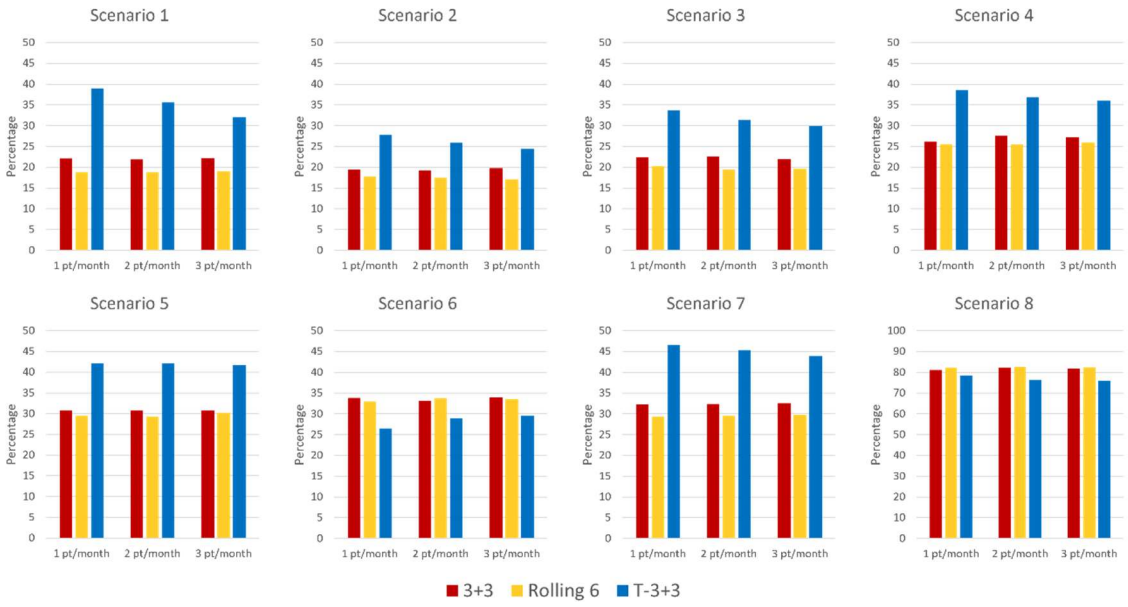


Figure 2.2: The correct MTD selection percentages of the 3+3, Rolling 6, and T-3+3 designs with different percentages of toxicity occurring in the latter half of the assessment window.

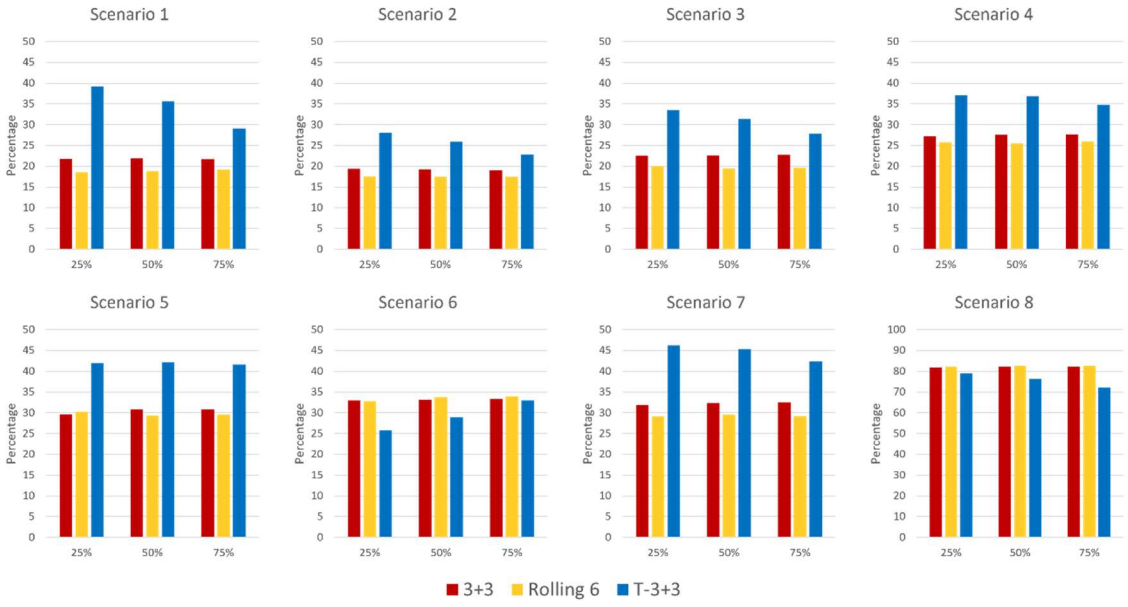
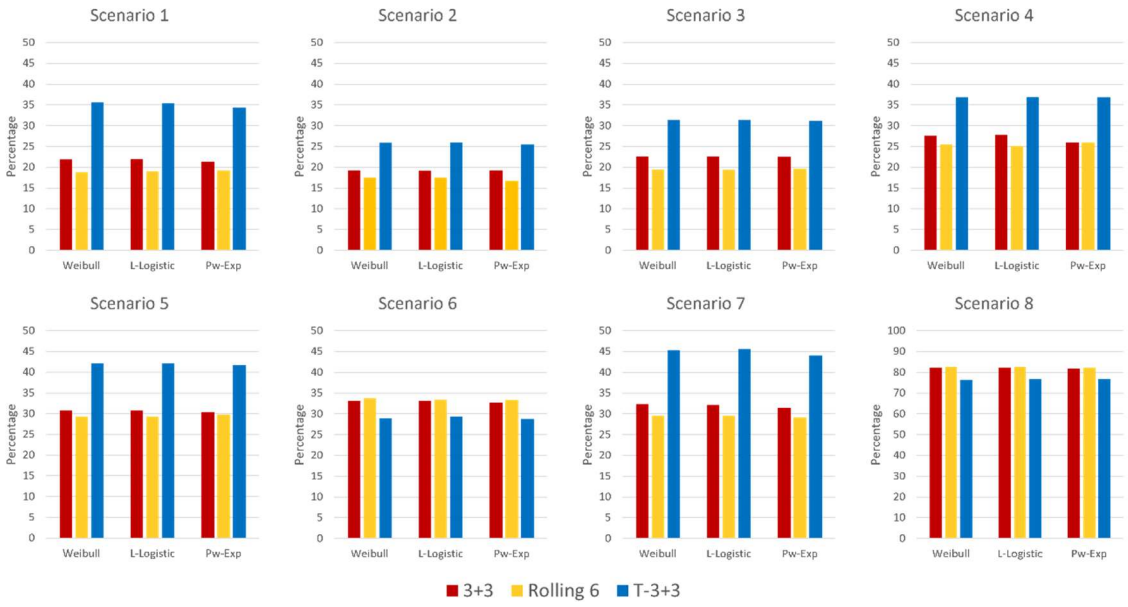


Figure 2.3: The correct MTD selection percentages of the 3+3, Rolling 6, and T-3+3 designs with different time-to-event distributions.



2.4 Trial Illustration

Using patient-level data in Table 2.3, Figure 2.4 provided a numerical trial comparison between the 3+3 and T-3+3 designs. We can see that the T-3+3 design significantly shortens trial duration while having similar decisions to the conventional 3+3 with trial suspension. We used the same trial setting where a new patient entered the trial every 15 days and was followed for 90 days for DLT assessment.

Using the T-3+3 design, on day 91, the first patient completed the follow-up without experiencing DLT. Based on decision rule 1 in Table 2.1, we escalated the dose and treated the second cohort of patients (patient ID 4-6) at dose 2. On day 171, patient 4 experienced DLT while patients 5 and 6 were still pending, resulting in AFR=0.75. Then, by applying decision rule 4, we suspended the trial for more information. On day 196, patient 5 also completed follow-up without experiencing DLT. Based on decision rule 3, we retained the current dose and treated 3 more patients (patient ID 7-9) at dose 2. On day 286, patient 7 completed follow-up without experiencing DLT. 6 patients were already treated at the current dose 2 with 1 DLT, 2 pending patients, and AFR=0.75. Based on decision rule 8, we escalated the dose and treated the next cohort of patients (patient ID 10-12) at dose 3. On day 376, patient 10 completed the follow-up without experiencing DLT. Based on decision rule 1 in Table 2.1, we escalated the dose and treated the next cohort of patients (patient ID 13-15) at dose 4. On day 446, patient 13 experienced DLT and AFR=0.53. So we suspended the trial based on decision rule 4. Then, on day 466, patient 15 also experienced DLT, and we de-escalated the dose based on decision rule 5 and treated the next cohort of patients (patient ID 16-18) back to dose 3. Then, dose 4

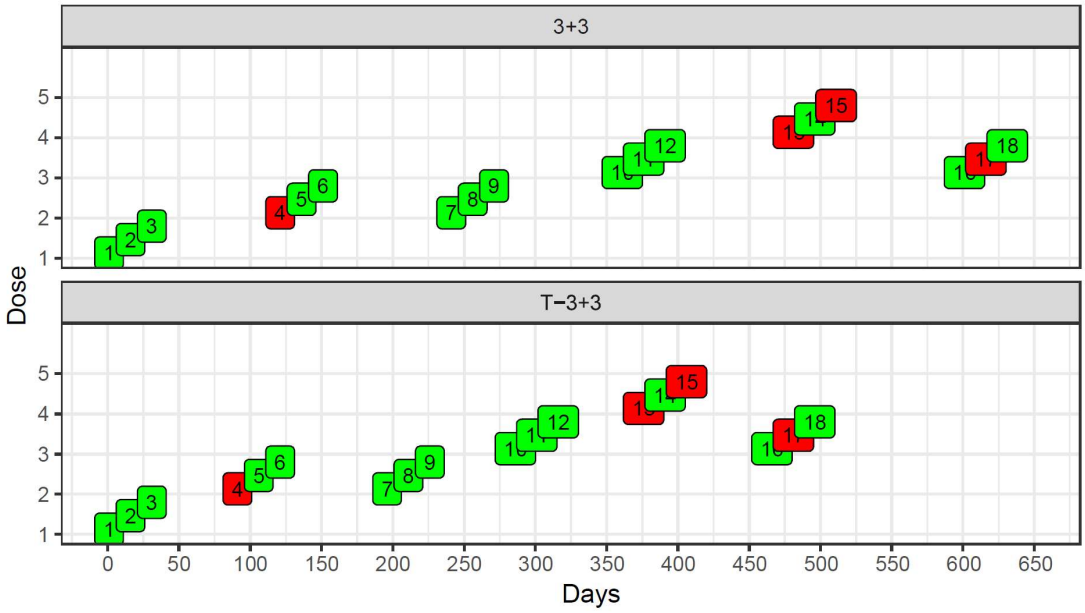
was overly toxic, with 2 of 3 patients experiencing DLT, so dose escalation was not allowed. Also, 6 patients had already been treated at doses 3 and 2, so dose retaining or de-escalation was also not allowed. Therefore, we completed the follow-up for the last cohort of patients and stopped the trial. After applying the isotonic regression method to the complete DLT data, we identified dose 3 as the MTD for a targeted rate of 0.3. The total trial duration was 586 days. As a comparison, if the conventional 3+3 design was used for the same dataset, we would need 721 days to reach the same conclusion.

Table 2.3: Hypothetical patient-level data for trial illustration

3 + 3					T - 3 + 3				
ID	Dose	DLT	Day	Time to DLT	ID	Dose	DLT	Day	Time to DLT
1	1	0	1		1	1	0	1	
2	1	0	16		2	1	0	16	
3	1	0	31		3	1	0	31	
4	2	1	121	80	4	2	1	91	80
5	2	0	136		5	2	0	106	
6	2	0	151		6	2	0	121	
7	2	0	241		7	2	0	196	
8	2	0	256		8	2	0	211	
9	2	0	271		9	2	0	226	
10	3	0	361		10	3	0	286	
11	3	0	376		11	3	0	301	
12	3	0	391		12	3	0	316	
13	4	1	481	70	13	4	1	376	70
14	4	0	496		14	4	0	391	
15	4	1	511	60	15	4	1	406	60
16	3	0	601		16	3	0	466	
17	3	1	616	60	17	3	1	481	60
18	3	0	631		18	3	0	496	

Abbreviations: Day, day of enrollment; ID, patient ID. "-" indicates that the patient does not experience DLT at the end of the follow-up.

Figure 2.4: Trial illustration by applying the 3+3 design and the T-3+3 design.



2.5 Software

The R codes for implementation are available at https://github.com/yongzang2020/T-3+3. We will also develop graphical user interface-based R Shiny Apps to facilitate the application of the proposed design. The Apps will be freely available at http://www.trialdesign.org.

2.6 Conclusion

We have proposed a new method to allow the algorithm-based designs to handle delayed toxicity outcomes. The T-3+3 design can be implemented simply with pre-tabulated decision rules like the rolling six design. However, the proposed T-3+3 design has higher accuracy in identifying the MTD, can further shorten the trial duration, and is more flexible in choosing the target DLT rate. Compared with the conventional 3+3 design, The T-3+3 design supports continuous accrual without sacrificing patient safety or the accuracy of identifying the MTD, thus providing a practical phase I clinical trial design to accelerate early phase drug development.

Chapter 3

T-i3+3: i3+3 Design with Delayed Outcomes

3.1 Introduction

As breakthroughs in drug development continues to burgeon the number of promising candidate drugs, it has become crucial to determine the correct Maximum Tolerated Dose (MTD) levels in early phase dose-finding trials. Two general design approaches exist for phase I dose-finding trials. Algorithm-based trial designs make dose escalation and de-escalation decisions based on a set of prespecified rules. Model-based designs use statistical modeling to assign dosing decisions based on the modeled dose-toxicity relationship.

Rule-based designs, such as the 3+3 design, are more popular in clinical practice due to their simplicity and transparency when compared to model-based designs. Despite being easy to implement and simple to understand, the 3+3 design is less efficient in identifying the true MTD (Yang *et al*., 2015). Recently, there has been a growing interest in a new class of designs known as model-assisted designs, including mTPI (Ji *et al.*, 2010) and mTPI-2 (Guo *et al*., 2017) designs, the cumulative cohort design (CCD) (Ivanova *et al*., 2007), and the Bayesian optimal interval design (BOIN) (Liu and Yuan, 2015). These designs offer comparable operating characteristics to classical model-based designs, while also having a simpler implementation similar to rule-based designs.

In 2020, Liu et al. (2020) proposed a model-assisted, yet rule-based, design called interval 3+3 (i3+3). The i3+3 design utilizes simple but comprehensive rules to account for

the variabilities in the observed data. Simulation studies have shown that the i3+3 design outperforms the traditional 3+3 design in terms of both identifying the true maximum tolerated dose (MTD) and ensuring trial safety. Furthermore, the i3+3 design exhibits similar operating characteristics to model-based phase I designs. Additionally, the i3+3 design is flexible as it can adapt to different target toxicity rate p_T values and cohort sizes (Liu *et al*., 2020).

Just like 3+3 design, i3+3 design is also susceptible to the delayed toxicity outcomes. Ignoring the delayed toxicity outcomes potentially exposes larger amounts of patients to overly toxic doses while underestimating the true toxicity probability. One approach would be to suspend the patient accrual and wait until fully observing patient's toxicity outcome then proceed to the next dose assignment. However, this may result in lengthening trial duration and incurring resource costs that would turn a promising trial into an unfeasible one.

We propose a new phase I trial design as the extension of the i3+3 design handling the delayed outcome issue, referred to as the T-i3+3 design. The T-i3+3 design share similar decision rules to the i3+3 design along with some of its the desirable features of simplicity and transparency. T-i3+3 decision rule can be pre-tabulated. This design does not require intensive computation and statistical consultation during trial conduct, which saves cost and trial resources. One of the most appealing advantages of the T-i3+3 over the i3+3 is the significant reduction in trial duration which can amount to huge resource saving when multiple drug candidates being assessed.

3.2 Methods

3.2.1 Overview of the i3+3 Design and the Delayed Outcome

We start with an overview of the i3+3 design to set the stage for T-i3+3. Before the start of trial design, investigators must determine the target toxicity rate p_T, the highest acceptable toxicity rate for the MTD. Additionally, investigator must also determine an equivalence interval (EI), defined as $[p_T - \varepsilon_1,\ p_T + \varepsilon_2]$ which doses that have toxicity probability within the EI will be considered as the MTD. The EI allows for some variabilities and flexibility in clinical trials.

Once the target toxicity rate, p_T, and equivalence interval (EI) have been specified for the trial, patients are enrolled into the trial and assigned dose levels in cohorts. Suppose n is the number of patients who have been treated at the current dose level, and r is the number of patients who have experienced DLTs at the current dose, calculate two values $\frac{r}{n}$ and $\frac{(r-1)}{n}$. The dosing assignment for i3+3 for the next dose level is as follows: If $\frac{r}{n}$ is below the EI, $\left(\frac{r}{n} < p_T - \varepsilon_1\right)$, escalate to the next higher dose level. If $\frac{r}{n}$ is inside the EI, $\left(p_T - \varepsilon_1 < \frac{r}{n} < p_T + \varepsilon_2\right)$, continue to enroll patients at the current dose. When $\frac{r}{n}$ is above the EI, $p_T + \varepsilon_2 < \frac{r}{n}$, we look at $\frac{(r-1)}{n}$. If $\frac{(r-1)}{n}$ is below the EI, $(\frac{(r-1)}{n} < p_T - \varepsilon_1)$, continue to enroll patients at the current dose. If $\frac{(r-1)}{n}$ is inside or above the EI, $(\frac{(r-1)}{n} \geq p_T - \varepsilon_1)$, de-escalate to the next lower dose level. When current dose is the highest dose level and $\frac{r}{n}$ is below the EI, $\left(\frac{r}{n} < p_T - \varepsilon_1\right)$, continue to enroll patients at the current dose.

When the current dose is the lowest dose level and $\frac{r}{n}$ is above the EI, $(p_T + \varepsilon_2 < \frac{r}{n})$, then also continue to enroll patients at the current dose.

To avoid exposing patients to excessive toxicity, two safety rules are added as safeguards. Safety rule 1 is to terminate the trial when, at the lowest dose, $P(p_1 > p_T|r,n) > \xi$, often, $\xi = 0.95$. When there is a decision to escalate patients from dose level d to $d+1$, Safety rule 2 is to exclude dose level $d+1$ and higher doses and de-escalate to dose level d, if $P(p_{d+1} > p_T|r,n) > \xi$, often, $\xi = 0.95$. The trial is complete when a prespecified sample size is reached or if safety rule 1 occurs. When the trial is complete, the MTD is identified by isotonic regression.

The problem of delayed outcomes for the i3+3 design is fundamentally the same as for other phase I clinical trials, which is described in section 2.2.1. Issues due to delayed outcomes arise when patients' enrollment rate into the trial occurs faster than the assessment window. This causes some patients to still be in the assessment window when the interim dose decision for the upcoming patients needs to be made. These patients may or may not experience DLTs in their remaining follow-up time, so their DLT scores are missing. This problem plagues the i3+3 design because the needed DLT information for all patients in the current dose is required to determine the next dose level.

3.2.2 Probability Model and Dose-finding Algorithm

Our strategy for the T-i3+3 design uses features from the i3+3 and T-3+3 designs. When delayed outcomes occur in the trial, there will be some risk of making the incorrect dosing decision. By utilizing the existing DLT information and the remaining follow-up

time from the pending patients, we can derive the probabilities of the escalation, de-escalation and retain events. Among these, we choose the maximum probability for all the possible events. Lastly, we determine if there is sufficient information to continue accrual patient without trial suspension by comparing probability with prespecified cut-off values.

Suppose that a total of n patients have been treated at the current dose. Among these n patients, n_1 patients have completed the DLT follow-up period, and r_1 of them have experienced DLT, with $n_2 = n - n_1$ patients remaining at risk of delayed outcomes. In the presence of delayed toxicity outcome, we define U as the fixed assessment window and v_i as the actual follow up time for the ith patient at risk, $i = 1 \dots n_2$. $U - v_i$ is the follow up time remaining. Then, $\bar{v}$, the average follow-up time for all at-risk patients in the current dose is $\bar{v} = \frac{\sum_{i=1}^{n_2} v_i}{n_2}$. The distribution for the number of patients who will experience DLT, defined as r_2, among n_2 patients can be modeled by the following Beta-binomial distribution (see Appendix A for detailed derivation):

$$r_2 \sim Beta - binomial\left(n_2, k, r_1 + 1, n_1 - r_1 + n_2 \frac{\bar{v}}{U} + 1\right) \quad (1)$$

Therefore, let $\mathcal{D} = (r_1, n_1, \bar{v})$denote the observed data. The probability of $r_2 = k$ given $\mathcal{D}$, $\pi(r_2 = k|\mathcal{D})$ can be expressed as

$$\pi(r_2 = k|\mathcal{D}) = \frac{n_2!}{k!(n_2-k)!} \frac{B\left(k+r_1+1, n_2-k+n_1-r_1+n_2\frac{\bar{v}}{U}+1\right)}{B\left(r_1+1, n_1-r_1+n_2\frac{\bar{v}}{U}+1\right)} \quad (2)$$

Where $B(\cdot)$denotes the beta function.

With (2), we can calculate the probability of escalation, $\pi(E|\mathcal{D})$, de-escalation, $\pi(D|\mathcal{D})$, and dose retention, $\pi(R|\mathcal{D})$, given the observed data, $\mathcal{D}$. The probability of dose

escalation given the observed data, $\mathcal{D}$, is the probability that $\frac{r_1+r_2}{n_1+n_2}$ is below the equivalence interval, i.e., less than $p_T - \varepsilon_1$.

$$\pi(E|\mathcal{D}) = \Pr\left(\frac{r_1 + r_2}{n_1 + n_2} < p_T - \varepsilon_1\right)$$

The probability of de-escalation given the observed data, $\mathcal{D}$, is the probability of $\frac{r_1+r_2}{n_1+n_2}$ being above the equivalence interval, i.e., greater than $p_T + \varepsilon_2$; and $\frac{r_1+r_2-1}{n_1+n_2}$ being inside or above the equivalence interval, i.e., greater than or equal to $p_T - \varepsilon_1$.

$$\pi(D|\mathcal{D}) = \Pr\left(\frac{r_1 + r_2}{n_1 + n_2} > p_T + \varepsilon_2 \ \ \& \ \ \frac{r_1 + r_2 - 1}{n_1 + n_2} \geq p_T - \varepsilon_1\right)$$

Lastly $\pi(R|\mathcal{D}) = 1 - \pi(E|\mathcal{D}) - \pi(D|\mathcal{D})$.

After calculating, $\pi(E|\mathcal{D})$, $\pi(R|\mathcal{D})$ and $\pi(D|\mathcal{D})$, we choose the action that yields the maximum probability. However, if the maximum probability is not large enough (compared with a cut-off value, i.e., ϕ_E, ϕ_R and ϕ_D), we should suspend the trial and wait for more information.

We define 3 cutoff values ϕ_E, ϕ_R and ϕ_D for necessary trial suspension in the same way as the T-3+3 design from section 2.2.2. If a dose escalation is recommended by comparing the event happening probabilities, we compare $\pi(E|\mathcal{D})$ with ϕ_E. We escalate the current dose if $\pi(E|\mathcal{D})$> ϕ_E. $\pi(E|\mathcal{D}) \leq \phi_E$ indicates suspension of the trial to for more follow up data. Similar comparisons will be made for dose de-scalation and dose retaining, with ϕ_D and ϕ_R respectively. If all patients in the current dose level are still at risk of pending outcomes at dosing assignment time, the trial should be suspended and wait for more follow up data to strengthen trial safety. Computer simulated studies help choose optimal values to recommend the balance of patient allocation, trial duration and

MTD selection. Based on our preliminary simulation studies, we recommend using $\phi_E = 0.5$, $\phi_R = 0.5$, and $\phi_D = 0.75$ as the default values.

The ability to pre-tabulate the dose escalation/de-escalation rule into the trial protocol before the commencement of the trial is a key feature that greatly simplifies design implementation. The key is to link the ratio between the averaged follow-up time for at-risk patients at current dose, $\bar{v}$, and the total assessment window U, which is $\frac{\bar{v}}{U}$ (averaged follow-up ratio (AFR)), with the probabilities of events happening. When the AFR is given, together with r_1, n_1, and n_2, which are all immediately observed, we can calculate each event happening probability and determine the dosing movement. Note that the decision rule of T-i3+3 design is invariant to the duration of the assessment window, because AFR is standardized by U. This gives us the convenience that same decision table can be used for dose escalation and de-escalation without regard to the length of the assessment window, when p_T and EI is specified.

Like the i3+3 design, we include two safety rules to avoid excessive toxicity as ethical safeguards. For safety rule 1, we terminate the trial due to excess toxicity if at the lowest dose, $P(p_1 > p_T|r_1, n) > \xi$, where $\xi = 0.95$, as all the doses under investigation are overly toxic. When there is a decision to escalate patients from dose level d to $d + 1$, safety rule 2 is to exclude dose level $d + 1$ and higher doses from the trial and continue to dose at dose level d, if $P(p_{d+1} > p_T|r_1, n) > \xi$, where $\xi = 0.95$.

The trial is terminated when a prespecified sample size is reached or when safety rule 1 is reached. At the end of the trial, all patients complete follow-up with DLT data. The final step is to select the MTD. We identify the MTD using isotonic regression method

described in section 2.2.3. Free software to calculate the isotonic estimates and select the MTD will be made available.

3.3 Computer simulations

3.3.1 Simulation Setup

We compared the T-i3+3, and i3+3 designs with computer simulations under 8 representative scenarios with different dose-toxicity curves. The target toxicity rates p_T for the MTD were 0.17 (EI: [0.12, 0.22]) and 0.3 (EI: [0.25, 0.35]). We simulated a phase I clinical trial with 6 doses and a DLT assessment window of 3 months. The patient arrival time was simulated from an exponential distribution with the accrual rate of 2 patients per month. The time-to-DLT outcomes were simulated from Weibull distribution by controlling that 50% of the DLT occurred in the latter half of the assessment window. We considered 5 performance metrics based on 10,000 simulated trials.

1. Percentage of correct selection (PCS) of the MTD.
2. Percentage of patients allocated to the MTD.
3. Average trial duration.
4. Average number of patients.
5. Average number of DLTs.

Metrics 1 and 2 measured the accuracy of the design; metrics 3 and 4 measured the cost effectiveness; and metric 5 measured the safety.

3.3.2 Simulation Results

Table 3. 1 summarized the computer simulation results with target toxicity rate of 0.17 under 8 representative scenarios with different dose-toxicity curves. The MTD selection percentage under the dose level "0" represented the probability of selecting early termination of trial for toxicity. The numbers in bold were PCS of the MTD for both study designs. From Table 3.1, the T-i3+3 design generally performed comparably with the i3+3 design in terms of metrics 1 (PCS), 2 (percentage of patients allocated to the MTD), 4 (average number of patients) and 5 (average number of DLTs). Notably, there was substantial improvement in T-i3+3 over the i3+3 across all 8 scenarios for average trial duration. Take scenario 1 as an example: Dose level 6 was assigned as the true MTD with a DLT rate of 0.16. The i3+3 design yielded a 36.9% of correct selection of MTD compared to T-i3+3 design at 44.0%, which was about 7.1% higher than the i3+3 design. The percentage of patients allocated to the MTD for the T-i3+3 was 23.5 VS 18.8 for the i3+3 design. The average numbers of patients and the average numbers of DLTs were similar for both designs (35.8 VS 35.9 and 4.0 VS 3.9, for T-i3+3 and i3+3 respectively). The T-i3+3 design had a significantly shorter trial duration of 31.6 months, which was about 21 months shorter than that of the i3+3 design. Take scenario 6 as another example: the lowest dose, level 1, was assigned as the true MTD with a DLT rate of 0.14. The T-i3+3 and i3+3 designs yielded similar correct selection of MTD at 48.9% VS 50.0%. The percentages of patients allocated to the MTD for the T-i3+3 and i3+3 design were 45.3% VS 51.5%. The average numbers of patients were similar for both designs (33.8 VS 33.4), as were the

average numbers of DLTs (7.6 VS 6.9) for T-i3+3 and i3+3. The trial duration of the T-i3+3 design was 28.2 months, which is 20.4 months shorter than that of i3+3 design.

We now examined computer simulation results of another commonly used target toxicity rate of 0.3 in clinical settings. Table 3.2 summarized simulated results under 8 scenarios with different dose-toxicity curves. The trial duration of the T-i3+3 design was much shorter than that of the i3+3 design in all 8 scenarios. Again, the T-i3+3 design performed comparably with the i3+3 design in other metrics, 1, 2, 4 and 5.

Additionally, we examined the robustness of the results observed above under different target toxicity rates. The two trial designs under 0.1, 0.3 and 0.4 target toxicity rates were simulated under 8 scenarios (Figures 3.1-3.2; Tables 6.1-6.2). The trial durations of the T-i3+3 design were between 30% to 40% shorter than those of the i3+3 design; the T-i3+3 design performed similarly with the i3+3 design in the correct selection percentage of the MTD and other metrics.

To compare the operating characteristics of the T-i3+3 VS i3+3 design under various settings, we held the target toxicity rate fixed at 0.3 and simulated the two designs under 25%, 50% and 75% percentage of toxicity occurring in the latter half of the assessment window, and under Weibull, Log-logistic and Piecewise-exponential time-to-toxicity distributions. For page limit reason, we presented comparison results for trial duration and percentage correct selection of the MTD in Figures 6.4 to 6.7. The results were consistent with above simulations. The trial durations of the T-i3+3 design were significantly shorter than those of the i3+3 design and the correct selection percentages of the MTD were comparable between the T-i3+3 design and the i3+3 design. This

confirmed robustness of the desirable operating characteristics of T-i3+3 compared to i3+3 under various conditions.

Table 3.1: The operating characteristics of the i3+3 and T-i3+3 designs, including percentages of MTD selection (Selection (%)) and patients' allocation (Patients (%)) at each dose level, number of patients in the trial and trial duration, based on 10,000 simulated trials. The targeted DLT rate is 0.17.

Designs		Dose level							Duration (month)	No. of Patients	No. of DLTs
		0	1	2	3	4	5	6			
					Scenario 1						
	True toxicity		0.04	0.08	0.10	0.12	0.14	**0.16**			
i3 + 3	Selection(%)	0.4	2.7	8.9	13.6	17.3	20.3	**36.9**	52.7	35.9	3.9
	Patients (%)		13.7	18.2	18.6	16.8	14.0	18.8			
T-i3 + 3	Selection(%)	0.5	2.7	6.5	10.2	15.4	20.7	**44.0**	31.6	35.8	4.0
	Patients (%)		13.2	15.9	16.6	16.0	14.8	23.5			
					Scenario 2						
	True toxicity		0.03	0.06	0.09	0.12	**0.15**	0.34			
i3 + 3	Selection(%)	0.2	1.4	5.2	13.3	21.8	**45.7**	12.4	52.8	36.0	4.6
	Patients (%)		11.6	15.6	19.0	19.2	22.2	12.4			
T-i3 + 3	Selection(%)	0.2	1.2	4.4	10.6	20.9	**49.2**	13.4	31.7	35.9	5.0
	Patients (%)		11.4	14.4	16.8	17.9	23.6	15.9			
					Scenario 3						
	True toxicity		0.03	0.05	0.10	**0.17**	0.30	0.44			
i3 + 3	Selection(%)	0.4	1.0	6.0	25.4	**46.5**	19.4	1.4	52.5	35.9	5.3
	Patients (%)		11.2	16.0	24.8	28.5	15.5	4.1			
T-i3 + 3	Selection(%)	0.1	0.8	5.3	23.9	**47.7**	20.2	2.0	31.1	36.0	5.9
	Patients (%)		10.9	14.6	22.2	27.6	18.0	6.6			
					Scenario 4						
	True toxicity		0.03	0.10	**0.15**	0.20	0.25	0.30			
i3 + 3	Selection(%)	0.3	4.0	20.3	**29.8**	25.2	14.7	5.7	52.6	35.9	5.2
	Patients (%)		14.6	25.3	25.9	19.0	10.2	5.1			
T-i3 + 3	Selection(%)	0.2	3.7	17.2	**26.8**	25.7	18.4	8.0	30.9	35.9	5.6
	Patients (%)		14.2	22.0	23.4	18.9	12.9	8.6			

Table 3.1 continues.

Designs		Dose level							Duration (month)	No. of Patients	No. of DLTs
		0	1	2	3	4	5	6			
					Scenario 5						
	True toxicity		0.08	**0.17**	0.25	0.33	0.41	0.49			
i3 + 3	Selection(%)	2.2	21.7	**44.7**	23.9	6.5	0.9	0.1	51.4	35.3	6.3
	Patients (%)		30.7	36.7	21.9	8.1	2.1	0.4			
T-i3 + 3	Selection(%)	1.7	19.4	**42.3**	26.4	8.4	1.8	0.1	29.2	35.5	6.9
	Patients (%)		26.9	33.4	23.6	11.1	4.0	1.0			
					Scenario 6						
	True toxicity		**0.14**	0.24	0.34	0.44	0.54	0.64			
i3 + 3	Selection(%)	9.0	**50.0**	33.7	6.7	0.7	0.1	0.0	48.6	33.4	6.9
	Patients (%)		51.5	33.5	11.9	2.6	0.4	0.0			
T-i3 + 3	Selection(%)	7.8	**48.9**	34.2	8.1	0.9	0.1	0.0	28.2	33.8	7.6
	Patients (%)		45.3	33.5	15.2	4.8	1.0	0.1			
					Scenario 7						
	True toxicity		0.02	0.04	0.06	0.08	0.10	**0.13**			
i3 + 3	Selection(%)	0.2	0.5	2.0	4.9	10.2	19.3	**63.0**	53.0	36.0	3.1
	Patients (%)		10.1	12.4	14.3	15.6	16.5	31.2			
T-i3 + 3	Selection(%)	0.0	0.3	1.4	4.4	8.8	17.5	**67.6**	31.9	36.0	3.1
	Patients (%)		9.9	11.7	13.4	14.3	15.6	35.2			
					Scenario 8						
	True toxicity		0.30	0.44	0.56	0.65	0.70	0.83			
i3 + 3	Selection(%)	67.5	31.8	0.7	0.0	0.0	0.0	0.0	29.8	20.8	6.7
	Patients (%)		85.1	13.3	1.5	0.1	0.0	0.0			
T-i3 + 3	Selection(%)	63.6	35.4	0.9	0.1	0.0	0.0	0.0	21.9	22.5	7.5
	Patients (%)		78.7	17.3	3.5	0.5	0.0	0.0			

Table 3.2: The operating characteristics of the i3+3 and T-i3+3 designs, including percentages of MTD selection (Selection (%)) and patients' allocation (Patients (%)) at each dose level, number of patients in the trial and trial duration, based on 10,000 simulated trials. The targeted DLT rate is 0.3.

Designs		Dose level							Duration (month)	No. of Patients	No. of DLTs
		0	1	2	3	4	5	6			
					Scenario 1						
	True toxicity		0.03	0.05	0.10	0.15	0.20	**0.32**			
i3 + 3	Selection(%)	0.0	0.0	0.3	1.9	9.0	32.7	**56.1**	52.6	36.0	6.4
	Patients (%)		9.3	10.4	13.7	17.7	22.5	26.4			
T-i3 + 3	Selection(%)	0.0	0.0	0.2	2.7	10.3	32.0	**54.8**	32.7	36.0	6.4
	Patients (%)		9.3	10.7	14.1	17.1	21.3	27.5			
					Scenario 2						
	True toxicity		0.06	0.12	0.18	0.24	**0.30**	0.36			
i3 + 3	Selection(%)	0.0	0.3	3.4	15.9	29.0	**29.2**	22.2	52.3	36.0	7.5
	Patients (%)		11.1	16.4	22.5	23.1	16.5	10.3			
T-i3 + 3	Selection(%)	0.0	0.3	3.6	16.7	28.1	**29.0**	22.4	32.1	36.0	7.6
	Patients (%)		11.5	16.6	21.6	21.3	16.4	12.7			
					Scenario 3						
	True toxicity		0.05	0.10	0.20	**0.30**	0.44	0.56			
i3 + 3	Selection(%)	0.0	0.2	3.5	28.3	**47.6**	18.5	1.9	52.1	36.0	8.5
	Patients (%)		10.4	16.3	28.7	29.3	12.9	2.5			
T-i3 + 3	Selection(%)	0.0	0.2	3.8	28.8	**46.4**	18.9	1.8	31.2	36.0	8.7
	Patients (%)		10.7	16.6	27.6	27.2	14.0	3.9			
					Scenario 4						
	True toxicity		0.10	0.21	**0.32**	0.49	0.56	0.68			
i3 + 3	Selection(%)	0.2	3.7	32.9	**51.3**	10.7	1.3	0.1	51.8	35.9	9.6
	Patients (%)		17.6	34.5	34.3	11.8	1.7	0.1			
T-i3 + 3	Selection(%)	0.1	3.6	33.9	**49.9**	11.0	1.5	0.1	29.6	36.0	9.8
	Patients (%)		18.3	32.7	32.7	13.1	2.8	0.4			

Table 3.2 continues.

Designs		Dose level 0	1	2	3	4	5	6	Duration (month)	No. of Patients	No. of DLTs
					Scenario 5						
	True toxicity		0.14	**0.30**	0.44	0.58	0.72	0.86			
i3 + 3	Selection(%)	0.7	18.4	**60.5**	18.8	1.5	0.0	0.0	51.4	35.8	10.3
	Patients (%)		31.2	45.5	19.6	3.5	0.2	0.0			
T-i3 + 3	Selection(%)	0.6	17.9	**61.2**	18.8	1.5	0.1	0.0	28.5	35.8	10.6
	Patients (%)		30.6	42.9	20.9	5.0	0.6	0.0			
					Scenario 6						
	True toxicity		**0.27**	0.36	0.45	0.54	0.63	0.72			
i3 + 3	Selection(%)	9.4	**47.3**	33.5	8.6	1.2	0.0	0.0	48.1	33.6	10.8
	Patients (%)		55.3	31.7	10.7	2.0	0.2	0.0			
T-i3 + 3	Selection(%)	8.3	**48.8**	32.5	9.0	1.3	0.2	0.0	28.0	34.0	11.2
	Patients (%)		53.3	30.6	12.3	3.2	0.6	0.1			
					Scenario 7						
	True toxicity		0.04	0.08	0.12	0.16	0.20	**0.24**			
i3 + 3	Selection(%)	0.0	0.1	0.5	3.5	11.0	21.3	**63.7**	52.7	36.0	5.8
	Patients (%)		9.8	12.0	15.3	17.7	18.1	27.1			
T-i3 + 3	Selection(%)	0.0	0.1	0.7	3.9	12.3	21.6	**61.4**	32.7	36.0	5.8
	Patients (%)		10.0	12.1	15.2	17.5	17.5	27.8			
					Scenario 8						
	True toxicity		0.42	0.55	0.63	0.70	0.78	0.85			
i3 + 3	Selection(%)	56.9	41.7	1.4	0.0	0.0	0.0	0.0	33.4	23.8	10.4
	Patients (%)		87.0	12.0	1.0	0.0	0.0	0.0			
T-i3 + 3	Selection(%)	53.1	45.0	1.8	0.1	0.0	0.0	0.0	23.1	25.6	11.3
	Patients (%)		82.7	14.7	2.4	0.3	0.0	0.0			

Figure 3.1: The trial duration in month of the i3+3 and T-i3+3 designs with different targeted DLT rate.

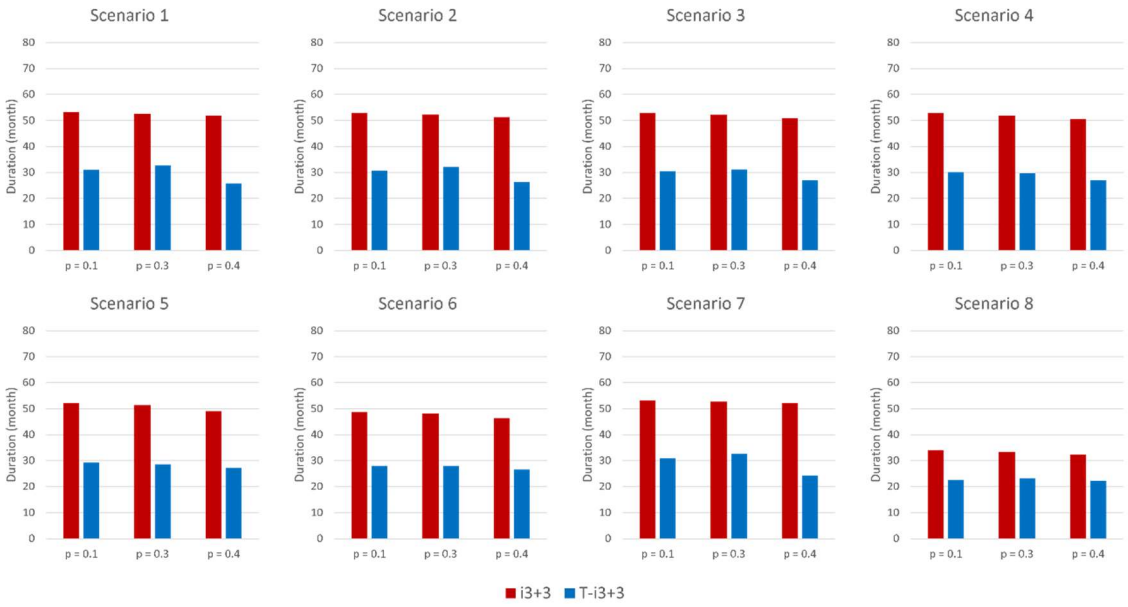
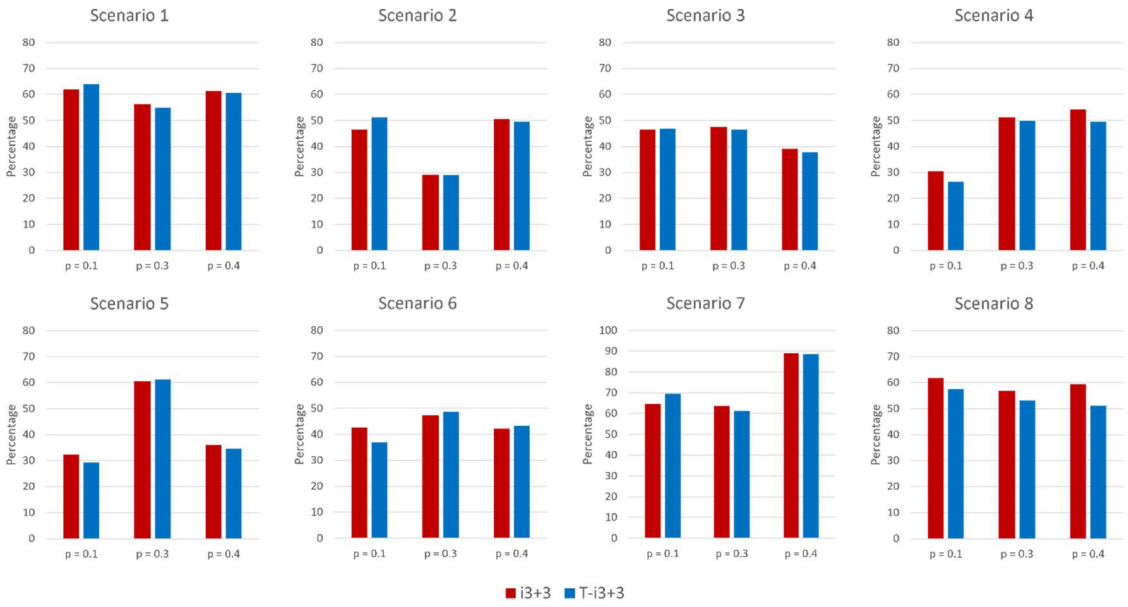


Figure 3.2: The correct MTD selection percentages of the i3+3 and T-i3+3 designs with different targeted DLT rate.



3.4 Conclusion

We have proposed a new model-assisted method, the T-i3+3 design, to handle delayed toxicity outcome. The T-i3+3 design is both easy to implement using pre-tabulated decision rules and statistically rigorous allowing for higher accuracy in identifying the correct MTD. In addition, the T-i3+3 design has built in safety features that prevent unnecessary patients from being exposed to overly toxic doses.

Moreover, in the presence of delayed outcomes, the T-i3+3 design significantly reduces trial duration compared to the conventional i3+3 design which often suspends the trial. Despite the shorter duration, the T-i3+3 design maintains comparable operating characteristics. Thus, T-i3+3 is a useful phase I dose finding trial design that can provide significant insight and resource savings to early phase drug development.

Chapter 4

An Adaptive Umbrella Trial Design Using Bayesian Latent Class Model

4.1 Introduction

Thanks to the improved understanding of cancer biology and rapid advancements in biomedicine, targeted therapies have become a prominent approach in clinical oncology (Sawyers, 2004; Simon and Roychowdhury, 2013; Lee *et al*., 2018). These therapies have led to a better understanding of cancer-causing mechanisms and the identification of corresponding therapeutic targets. However, due to the heterogeneous biological profiles of patients, specific targeted agents may not be effective for the general population of patients. Instead, they may only benefit a small proportion of patients or not work at all. To address this issue, the clinical application of targeted agents requires the identification of predictive biomarkers that can be used to identify subgroups of patients who are likely to benefit from the targeted agents (Mandrekar and Sargent, 2009; Freidlin *et al*., 2010; Sargent and Mandrekar, 2013). This personalized approach allows for more effective treatment and better outcomes for patients.

The umbrella trial is a clinical trial strategy that accommodates the paradigm shift towards personalized medicine (Woodcock and LaVange, 2017; Dean *et al*., 2020). Unlike the traditional one-drug-one-indication approach, the umbrella trial evaluates multiple investigational drugs in different subgroups of patients with the same disease. These subgroups are defined by various biomarkers, such as genetic or molecular aberrations. As shown in Figure 4.1, patients with the same disease are enrolled in the trial and then

classified into putatively heterogeneous subgroups based on their biomarker profiles. Within each subgroup, patients are randomized to multiple investigational treatment arms and a control arm for subgroup-specific treatment evaluation. The umbrella trial evaluates the effects of multiple targeted therapies on different subgroups of patients under a unified master protocol and is an efficient strategy to optimize clinical findings in a timely and safe manner (Simon and Roychowdhury, 2013; Berry, 2015). Due to its flexibility and efficiency, the umbrella trial has drawn tremendous attention in the clinical community (Zhou *et al*., 2008; Barker *et al*., 2009; Printz, 2010; Kim *et al*., 2011; Berry *et al*., 2012; Mullard, 2015; Lee *et al*., 2019).

Compared to conventional cytotoxic agents like chemotherapy and radiotherapy, targeted therapies often do not achieve rapid tumor shrinkage. Instead, they frequently delay cancer progression and extend survival, with a significant number of patients experiencing long-term durable responses (Kroeger *et al*., 2014; Rao *et al*., 2018). Given this characteristic, progression-free survival (PFS) is a preferred efficacy endpoint in many phase II umbrella trials, rather than the conventional binary tumor response. Therefore, in this paper, we focus on PFS as the primary efficacy endpoint.

An essential aspect of umbrella trials is the use of patients' biomarker profiles to define subgroups and predict treatment effects. Typically, umbrella trials employ gene mutations or oncologic pathway activations as biomarkers, which are specific to the investigational agents. Targeted therapies are designed to modulate specific biomarkers in cancer cells to inhibit tumor growth. Consequently, the treatment effect of a targeted agent is typically highly dependent on the biomarker status that the agent targets. One

innovation of our design is to exploit this dependency to enhance the effectiveness of the umbrella trial.

Our study is motivated by an ongoing randomized controlled phase II umbrella trial in the Indiana University Simon Comprehensive Cancer Center (IUSCCC). This trial evaluates the efficacy of two novel targeted agents for patients with advanced non-small-cell lung cancer (NSCLC), based on their epidermal growth factor receptor (EGFR) and Kirsten rat sarcoma virus (KRAS) biomarker statuses. Patients can be classified into four subgroups (E-K-, E+K-, E-K+, E+K+) according to their EGFR and KRAS statuses, with each subgroup receiving either an EGFR-specific targeted agent, a KRAS-specific targeted agent, or a control chemotherapy Docetaxel. The primary efficacy endpoint is 12-month progression-free survival (PFS), and the clinical benefit rate (CBR) will also be recorded using the response evaluation criteria in solid tumors (RECIST 1.1). The goal of this umbrella trial is to match each subgroup with the most effective treatment. To achieve this goal, we propose a "we-learn-as-we-go" Bayesian adaptive design given the limited understanding of the treatment and subgroup properties.

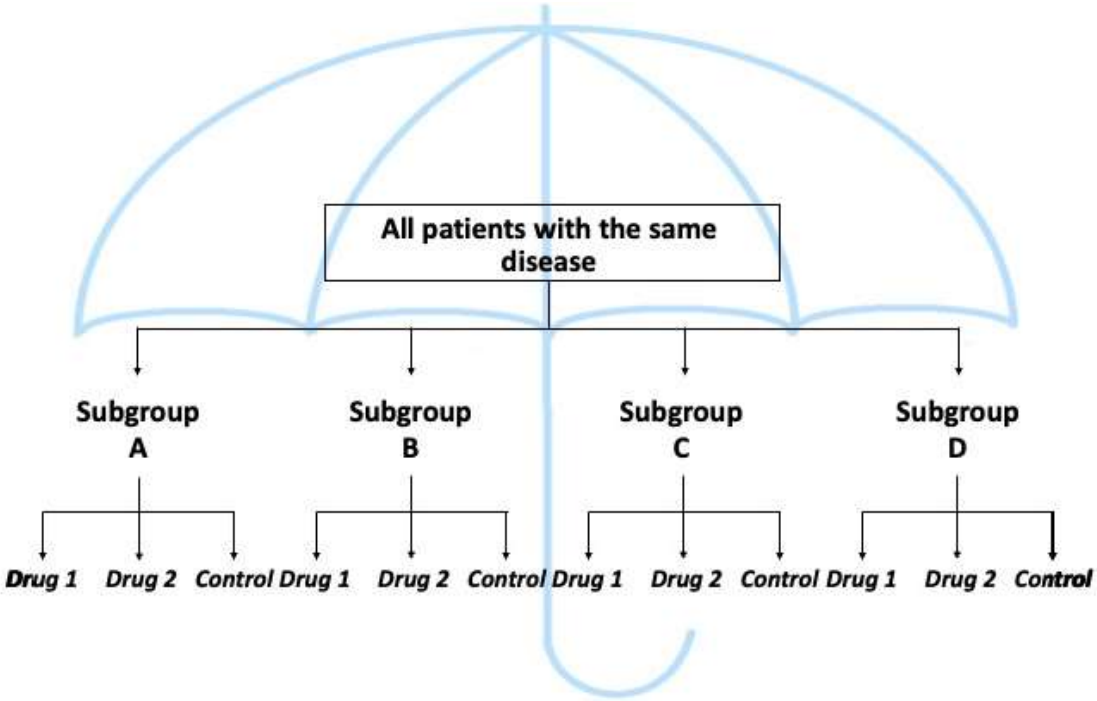
There are two main challenges in designing this motivating trial. The first challenge is that the use of PFS as the primary outcome requires extended follow-up time, which can impede the implementation of interim analysis. The second challenge is that slow accrual rates make it difficult to obtain enough sample size to detect subgroup-specific treatment effects if each subgroup is analyzed separately. This will reduce the statistical power to identify effective treatments.

To address these challenges, we propose a Bayesian adaptive umbrella trial design. To handle the first challenge, we suggest using the short-term CBR, which can be quickly ascertained, as a surrogate marker for the long-term PFS. We will develop a mixture regression model to borrow information between the CBR and PFS. To address the second challenge, we will use a data-driven Bayesian latent class model. This model will adaptively group subgroups into latent classes based on overall data heterogeneities. Response outcomes will then be pooled within the latent classes to overcome subgroup-specific sample size limitations.

Numerous umbrella trial designs have been proposed in the literature (Pusztai *et al*., 2007; Jones and Holmgren, 2007; Wason *et al*., 2015; Parashar *et al*., 2017; Dutton and Holmes, 2018; Chu and Yuan, 2018a, 2018b; Zang *et al*., 2019; Ballarini *et al*., 2021; Kang *et al*., 2021; Ren *et al*., 2021; Ouma *et al*., 2021; Zhao *et al*., 2023; Shan *et al*., 2023). However, few of these designs focus on survival outcomes, and information borrowing is mostly for treatment effects through the hierarchical model, rather than for biomarker-based subgroups through the latent class model, which are the novelties of our proposed research. For a systematic review on umbrella trial designs, see Ouma et al. (2022).

The remainder of this chapter is organized as follows. In Section 4.2, we describe the probability models for the short-term CBR and long-term PFS and the latent class modeling approach. In Section 4.3, we propose the Bayesian response-adaptive design using the developed models. In Section 4.4, we examine the operating characteristics of the proposed design through simulation studies. Finally, we provide concluding remarks in Section 4.5.

Figure 4.1: An illustrative example of the umbrella trial



4.2 Methods

4.2.1 Probability Model

Consider a randomized controlled umbrella trial designed to evaluate the treatment effects of multiple targeted agents in patients with the same disease but carrying different genetic or molecular aberrations. For ease of presentation, we assume that the patients' genetic or molecular aberrations can be represented by two biomarkers, B_1 and B_2, with $B_1 = 1$ indicating the presence of a mutation on B_1 and $B_1 = 0$ indicating its absence. The definition of B_2 follows the same pattern as B_1. Consequently, the patient population can be classified into four subgroups: $g = 0$ for $(B_1 = 0,\ B_2 = 0)$; $g = 1$ for $(B_1 = 1,\ B_2 = 0)$; $g = 2$ for $(B_1 = 0,\ B_2 = 1)$; and $g = 3$ for $(B_1 = 1,\ B_2 = 1)$.

We denote the treatment under investigation as t, where $t = 0$ represents the control treatment. In this study, we consider two targeted agents that match the number of biomarkers. We use $t = 1$ to represent the targeted agent designated for biomarker B_1, and $t = 2$ for the targeted agent designated for biomarker B_2. Based on the biomedical mechanism of the targeted agents, we speculate that $t = 1$ will have a better treatment effect in patients with $B_1 = 1$, while $t = 2$ will have a better effect in patients with $B_2 = 1$. The objective of the umbrella trial is to identify the effective subgroup-specific treatments for each subgroup g based on the PFS outcomes.

Let u_{gtl} denote the time to disease progression and let c_{gtl} denote the right-censoring time for the lth patient in subgroup g and receiving treatment t, $g = 0,1,2,3$;

$t = 0,1,2;\ l = 1, \dots, n_{gt}$. We observe $y_{gtl} = \min(u_{gtl}, c_{gtl})$ and the censoring indicator $\delta_{gtl} = I(u_{gtl} \le c_{gtl})$. Let x_{gtl} be the short-term efficacy response with $x_{gtl} = 1$ representing at least stable disease (CR/PR/SD) and $x_{gtl} = 0$ otherwise. Under the proportional hazards assumption for x_{gtl}, we model the hazard of the lth patient in subgroup g and receiving treatment t as follows:

$$\lambda(y|x_{gtl}, \beta_{gt}) = \lambda_{0gt}(y)\exp(\beta_{gt}x_{gtl}),$$

with $\lambda_{0gt}(y)$ standing for the baseline hazard function and β_{gt} standing for the logarithm of the hazard ratio for the short-term efficacy response with given subgroup g and treatment t. We take a parametric Weibull distribution for the baseline hazard function, $\lambda_{0gt}(y) = \left(\frac{\alpha_{gt}}{\kappa_{gt}}\right)\left(\frac{y}{\kappa_{gt}}\right)^{\alpha_{gt}-1}$. The corresponding survival function and density function are then given as

$$S(y|x_{gtl}, \alpha_{gt}, \kappa_{gt}, \beta_{gt}) = \exp\left(-exp(\beta_{gt}x_{gtl})\left(\frac{y}{\kappa_{gt}}\right)^{\alpha_{gt}}\right);$$

$$f(y|x_{gtl}, \alpha_{gt}, \kappa_{gt}, \beta_{gt}) = exp(\beta_{gt}x_{gtl})\left(\frac{\alpha_{gt}}{\kappa_{gt}}\right)\left(\frac{y}{\kappa_{gt}}\right)^{\alpha_{tg}-1}\exp\left(-exp(\beta_{gt}x_{gtl})\left(\frac{y}{\kappa_{gt}}\right)^{\alpha_{gt}}\right).$$

We further define $z_{gt} = \sum_{l=1}^{n_{gt}} x_{gtl}$, which follows a Binomial distribution $Bin(n_{gt}, p_{gt})$ with p_{gt} standing for the clinical benefit rate (CBR). Let $\mathcal{D} = \{(y_{gtl}, x_{gtl}, \delta_{gtl}), l = 1, \dots, n_{gt}; g = 0,1,2,3; t = 0,1,2\}$ be the observed data, and let $\theta = \{(p_{gt}, \alpha_{gt}, \kappa_{gt}, \beta_{gt}),\ g = 0,1,2,3; t = 0,1,2\}$ be the parameters of interest. Then, the overall likelihood function can be written as

$$L(\mathcal{D}|\theta) = \prod_{g=0}^{3}\prod_{t=0}^{2}(p_{gt}{}^{z_{gt}}\{1-p_{gt}\}^{n_{gt}-z_{gt}}\prod_{l=1}^{n_{gt}}\{f(y|x_{gtl},\theta)^{\delta_{gtl}}S(y|x_{gtl},\theta)^{1-\delta_{gtl}}\}) \quad (3)$$

If we treat all the θ's as independent, we can use the likelihood function (3) for statistical inference. However, this conventional method ignores the underlying biomarker profiles and the biomedical mechanism of the targeted agents, making it less powerful. To address this, we introduce a Bayesian latent class model that adaptively borrows information between the subgroups g and treatments t, improving the efficiency of the umbrella trial.

We begin by examining the case where $t = 0$, which serves as the standard control treatment. In this case, the treatment effects are typically constant across all subgroups, so we can set $p_{g0} = p_0$ and $\beta_{g0} = \beta_0$ for any subgroup g. Additionally, if a patient experiences $x_{gtl} = 0$, that indicates a significant tumor progression, which is a clinically severe adverse event and a strong signal for treatment failure. Typically, this leads to a significantly shortened PFS, regardless of the patient's biomarker profile. Given $x_{gtl} = 0$, the PFS distribution is solely determined by α_{gt} and κ_{gt}. It is therefore reasonable to assume $\alpha_{gt} = \alpha_t$ and $\kappa_{gt} = \kappa_t$ for any subgroup g.

The next step is to categorize the subgroup g into latent classes based on treatment $t = 1$ and $t = 2$. We define $\mathcal{M}_1(g)$ as the latent class index for treatment $t = 1$. The treatment $t = 1$ specifically targets for the genetic or molecular aberration on biomarker B_1. Therefore, it should yield homogeneous treatment effect for the subgroups of patients with $B_1 = 1$, regardless of the status of B_2. Therefore, we can always combine subgroups $g = 1$ $(B_1 = 1,\ B_2 = 0)$ and $g = 3$ $(B_1 = 1,\ B_2 = 1)$ as $g = 1 \cup g = 3$ for patients receiving $t = 1$. Then, if B_2 is also a predictive marker for $t = 1$ in the absence of B_1 mutation, subgroups $g = 0$ $(B_1 = 0,\ B_2 = 0)$ and $g = 2$ $(B_1 = 0,\ B_2 = 1)$

should be distinct subgroups, resulting in the following latent classes: $(\{g = 0\}, \{g = 2\}, \{g = 1 \cup g = 3\})$, denoted by $\mathcal{M}_1(g) = 1$. On the other side, if B_2 has no predictive marker effect on $t = 1$, subgroups $g = 0$ and $g = 2$ should be combined, resulting in the following latent classes: $(\{g = 0 \cup g = 2\},\ \{g = 1 \cup g = 3\})$, denoted by $\mathcal{M}_1(g) = 2$. Lastly, if biomarker B_1 has no predictive marker effect on $t = 1$, then the population is homogeneous with only one latent class $(\{g = 0 \cup g = 1 \cup g = 2 \cup g = 3\})$. Similarly, for treatment $t = 2$, we define $\mathcal{M}_2(g)$ as the latent class index and have the following latent class patterns: $(\{g = 0\}, \{g = 1\}, \{g = 2 \cup g = 3\})$ denoted by $\mathcal{M}_2(g) = 1$; $(\{g = 0 \cup g = 1\},\ \{g = 2 \cup g = 3\})$ denoted by $\mathcal{M}_2(g) = 2$; and class $(\{g = 0 \cup g = 1 \cup g = 2 \cup g = 3\})$ denoted by $\mathcal{M}_2(g) = 3$.

We use the latent class approach to determine the degree of information sharing among the parameters. Specifically, for any subgroups of the patients that receive the same treatment and belong to the same latent class, their corresponding p_{gt} and β_{gt} should be identical across g. For example, if we assume that $\mathcal{M}_1(g) = 2$ and $\mathcal{M}_2(g) = 3$ are the true latent class patterns, then we have the following parameters configuration: $(p_{01} = p_{21}, p_{11} = p_{31})$; $(\beta_{01} = \beta_{21}, \beta_{11} = \beta_{31})$; $(p_{02} = p_{12} = p_{22} = p_{32})$; $(\beta_{02} = \beta_{12} = \beta_{22} = \beta_{32})$.

In practice, however, the true latent class patterns are unknown. We propose to estimate the underlying patterns through the data. Specifically, we define S as the set of all possible pairs of latent class patterns $\{\mathcal{M}_1(g), \mathcal{M}_2(g)\}$, denoted as $S = (\{l_1, l_2\}, l_1 = 1,2,3;\ l_2 = 1,2,3)$ and assign $(\mathcal{M}_1(g), \mathcal{M}_2(g))$ the uniform categorical distribution $Cat(S)$ with $\Pr(\mathcal{M}_1(g) = l_1, \mathcal{M}_2(g) = l_2) = 1/9$. We also assign a non-informative beta

prior $Beta(0.5,0.5)$ for p_{gt}, a non-informative gamma prior $Gamma(0.001,0.001)$ for α_{gt} and κ_{gt}, and a non-informative normal prior $N(0, 10^2)$ for β_{gt}. Let $f(\theta)$ be the prior distributions for θ and $f(\theta, \mathcal{M}_1(g), \mathcal{M}_2(g)|\mathcal{D})$ be the posterior distributions for $(\theta, \mathcal{M}_1(g), \mathcal{M}_2(g))$, we have the following expression

$$f(\theta, \mathcal{M}_1(g), \mathcal{M}_2(g)|\mathcal{D}) \propto \left(\prod_{l_1,l_2=1}^{3} L\left(\mathcal{D}\middle|\theta, \mathcal{M}_1(g), \mathcal{M}_2(g)\right)^{I(\mathcal{M}_1(g)=l_1, \mathcal{M}_2(g)=l_2)}\right) f(\theta) \quad (4)$$

We propose to use the Bayesian data augmentation (BDA) method to derive the posterior distributions. The BDA consists of an imputation step (I) and a posterior sampling step (P), and can be summarized as follows:

1. Initiate starting values for θ.
2. I step. Impute $(\mathcal{M}_1(g), \mathcal{M}_2(g))$ from the categorical distribution $Cat(S)$ with the probabilities

$$\Pr(\mathcal{M}_1(g) = l_1, \mathcal{M}_2(g) = l_2|\mathcal{D}) = \frac{L\left(\mathcal{D}\middle|\theta, \mathcal{M}_1(g) = l_1, \mathcal{M}_2(g) = l_2\right)}{\sum_{i,j=1}^{3} L\left(\mathcal{D}\middle|\theta, \mathcal{M}_1(g) = i, \mathcal{M}_2(g) = j\right)}.$$

3. Given the imputed values of $(\mathcal{M}_1(g), \mathcal{M}_2(g))$, use the Metropolis-within-Gibbs-sampler algorithm to sequentially draw of the posterior samples of θ from $f(\theta|\mathcal{D}, \mathcal{M}_1(g), \mathcal{M}_2(g)) \propto L\left(\mathcal{D}\middle|\theta, \mathcal{M}_1(g), \mathcal{M}_2(g)\right) f(\theta)$.
4. Repeat steps 2 and 3 until the Markov chain converges.

The posterior samples of θ will be used to estimate the PFS and guide the trial conduct, which will be illustrated in the next section.

4.2.2 Adaptive Design

Given the proposed regression model, the mean expectation of the PFS with fixed g and t can be derived as

$$\mu_{gt} = (1 - p_{gt})\kappa_{gt}\Gamma\left(1 + \frac{1}{\alpha_{gt}}\right) + p_{gt}\kappa_{gt}\exp\left(-\frac{\beta_{gt}}{\alpha_{gt}}\right)\Gamma\left(1 + \frac{1}{\alpha_{gt}}\right),$$

with $\Gamma()$ denoting the Gamma function. The objective of the umbrella design is to test whether the targeted agents are more effective than the control treatment in each of the subgroup:

$$H_0: \mu_{gt} < \mu_{g0} + \tau \quad versus \quad H_1: \mu_{gt} \geq \mu_{g0} + \tau, \quad for\ t = 1,2.$$

Here $\tau \geq 0$ is the margin representing the minimal meaningful difference of the mean PFS, which is pre-specified by the physicians before the trial.

Under the Bayesian framework, by using the proposed likelihood function (3), posterior distribution (4) and the BDA algorithm, we can calculate the posterior probabilities

$$Q_{gt} = \Pr\left(\mu_{gt} \geq \mu_{g0} + \tau \middle| \mathcal{D}\right) = \int I\left(\mu_{gt} \geq \mu_{g0} + \tau\right) f(\theta|\mathcal{D}) d\theta, \quad for\ t = 1,2.$$

We will employ Q_{gt} to carry out the trial and identify the effective treatment for each subgroup. Since this is an adaptive design, we will perform interim analyses after every cohort of L patients is enrolled in the trial and their short-term CBR is available. Patients whose PFS outcomes are still missing at the time of interim analysis will be considered administratively censored.

To address the issue of sparse data at the beginning of the trial, we propose a two-stage randomization scheme. In Stage I, we will randomly assign each cohort of patients

within each subgroup $g = 0,1,2,3$ to receive one of the three treatments $t = 0,1,2$ using equal randomization. Stage I will consist of a total of N_1 cohorts of patients. No interim analysis will be conducted during Stage I, as its purpose is solely to collect data to facilitate response-adaptive randomization in Stage II.

In stage II, for each subgroup, we allow the early graduation for the evaluation of any targeted agents, due to either superiority or futility. In particular, we define ζ_l and ζ_u as the lower-bound and upper-bound cut-off values. Then, we can monitor the trial and make the go-no go decision for each subgroup by comparing the posterior probabilities $Q_{gt} = \Pr(\mu_{gt} \geq \mu_{g0} + \tau | \mathcal{D})$ with these two cut-off values during each interim analysis. If $l-1$th cohort of patients have been enrolled in the trial. We assign treatments to the lth cohort of patients using the following adaptive randomization scheme:

1. Based on the accumulated data $\mathcal{D}_{l-1}$, we update the posterior probabilities Q_{gt} for $t = 1,2$ and $g = 0,1,2,3$.
2. For any subgroup g and $t = 1,2$, if $Q_{gt} > \zeta_u$, we graduate the treatment t for superiority; if $Q_{gt} < \zeta_l$, we graduate the treatment t for futility. If both targeted agents are graduated, the whole subgroup is early terminated. Otherwise, let $\mathcal{C}_g$ be the treatments remaining in the subgroup g (including the control treatment) and define $\widehat{\mu_{gt}} = \int \mu_{gt} f(\theta|\mathcal{D}) d\theta$ as the posterior expectation of μ_{gt}, we randomize each new coming patient belonging to the subgroup g to the treatment $t \in \mathcal{C}_g$ with the probability $\frac{\widehat{\mu_{gt}}}{\sum_{t \in \mathcal{C}_g} \widehat{\mu_{gt}}}$.

3. We collect the data for the lth cohort of patients and update the accumulated data from $\mathcal{D}_{\mathrm{l}-1}$ to $\mathcal{D}_{\mathrm{l}}$.
4. We repeat steps 1-3 until either all the subgroups are early terminated or the maximum number of cohorts for stage II, N_2, is reached.

If the maximum number of cohorts is reached, we need to make a final treatment evaluation at the end of the trial. For this purpose, we define ζ_f as the cut-off value for the final analysis. Then for every subgroup g, if there are any treatment targeted agent t satisfying $Q_{gt} > \zeta_f$, we should claim the subgroup-specific treatment effectiveness of that targeted agent for subgroup g. Otherwise, if we have $Q_{gt} < \zeta_f$ for both $t = 1$ and 2, then we should recommend the control treatment $t = 0$. As a Bayesian design, the cutoff values ζ_l, ζ_u and ζ_f are tunning parameters of the design and are determined through simulation studies to achieve desirable operating characteristics. Based on our preliminary simulation studies, we recommend $\zeta_u = 0.99$; $\zeta_l = 0.05$; $\zeta_f = 0.85$ as the default values. As the proposed design is the **u**mbrella design using the **l**atent **c**lass **m**odel, we refer to it as the ULCM design hereafter.

4.3 Computer Simulations

4.3.1 Simulation Setup

We conducted comprehensive simulation studies to investigate the operating characteristics (OCs) of the proposed ULCM design. A prevalence of 40% for B_1 mutation and 50% for B_2 mutation was specified, and a fixed follow-up time of 12 months was set for each patient. The margin τ was set to 0. In stage I, we randomly assigned 250 patients

(25 cohorts with 10 patients in each cohort) equally to the 3 treatments within each subgroup. In stage II, we adaptively randomized an additional 250 patients using the randomization scheme proposed in Section 3. PFS outcomes were generated from a piece-wise exponential distribution, with hazard rates determined by survival rates at the end of 12 months.

We compared the ULCM design with two alternative conventional designs. The first design was similar as the ULCM design but treated each subgroup of patients defined by B_1 and B_2 independently and evaluated the subgroup-specific treatment effects separately without using the proposed latent class model. The second design was more conventional and used the PFS outcomes only without the information borrowing from the short-term response outcome. We referred to these two designs as the Conv 1 and Conv 2 designs respectively.

We considered six different treatment-outcome scenarios for simulation. The following 3 performance metrics were used to evaluate and compare the 3 designs under investigation based on 5000 simulated trials:

1. percentage of correct treatment selection.
2. mean number of patients allocated to each treatment within each subgroup.
3. mean sample size.

Metric 1 measured the accuracy of the treatment evaluation, metric 2 measured the trial ethics and metric 3 measured the cost-effectiveness.

4.3.2 Simulation Results

Table 4.1 summarized the results. The true CBR p_{gt} and 12 months PFS rate defined by $\bar{F}_{gt}$ for each subgroup $g = 0,1,2,3$ and treatment $t = 0,1,2$ were also listed at the beginning of each scenario. The correct treatment selection percentages were indicated in bold fonts.

Scenario 1 evaluates the case where treatments 1 and 2 had the same effectiveness in all subgroups. The ULCM design achieved a correct treatment selection percentage of around 97% for both treatments 1 and 2 in all subgroups, which was about 10%~15% higher than the Conv 1 design without the latent class model and about 20%~30% higher than the Conv 2 design without either the latent class model or the joint modeling of short-term and long-term outcomes. The ULCM design also reduced the number of patients needed by about 60 and 90 compared to the Conv 1 and Conv 2 designs, respectively. Regarding patients' allocation, all designs performed similarly.

Scenario 2 represented a null situation where neither treatment 1 nor treatment 2 had a better PFS rate than the control. The ULCM design showed approximately 82% accuracy in claiming treatment ineffectiveness for both treatment 1 and treatment 2 in all subgroups, which again outperformed the Conv 1 and Conv 2 designs across all subgroups, and the improvement can be as large as 13% in subgroup $g = 1$ compared to the Conv 1 design. The ULCM design also had the smallest mean sample size, although the differences were moderate (around 10 patients).

In scenario 3, treatment 1 was effective only for patients containing the B_1 mutation (PFS rate 0.6 for $g = 1,3$), and treatment 2 was effective only for patients

containing the B_2 mutation (PFS rate 0.6 for $g = 2,3$). For subgroup $g = 0$ indicating $\{B_1 = 0,\ B_2 = 0\}$, the ULCM design had a correct treatment selection percentage of 78.2% for the control $t = 0$, which was similar to the Conv 1 and Conv 2 designs with percentages of 76.2% and 75.4%, respectively. For subgroup $g = 1$ indicating $\{B_1 = 1,\ B_2 = 0\}$, the ULCM design had a correct selection percentage of 83.2% for selecting treatment 1 only, which was about 10% higher than the Conv 1 design and 20% higher than the Conv 2 design. For subgroup $g = 2$ indicating $\{B_1 = 0, B_2 = 1\}$, the ULCM design had a correct selection percentage of 85.0% for selecting treatment 2 only. The Conv 1 design has a similar correct selection percentage of 81.8%, while the Conv 2 design has a much lower correct selection percentage of 70.2%. For the last subgroup $g = 3$ indicating $\{B_1 = 1,\ B_2 = 1\}$, the ULCM design had a correct selection percentage of 93.8% for selecting both treatments 1 and 2, which was about 8% higher than the Conv 1 design and 10% higher than the Conv 2 design.

In Scenario 4, treatment 1 was completely ineffective while treatment 2 was only effective for patients with the B_2 mutation. In Scenario 5, treatment 2 was more effective than treatment 1 for patients with the B_2 mutation, while treatment 1 was more effective for patients with the B_1 mutation. In Scenario 6, treatment 1 was equally effective for all patients, while treatment 2 was only effective for patients with the B_2 mutation. Results for Scenarios 4 to 6 were consistent with Scenarios 1 to 3. The ULCM design consistently achieved the highest correct treatment selection percentage across all subgroups and scenarios and always had the smallest mean sample sizes. Patients' allocation performance was similar across all designs.

Although response adaptive randomization had been utilized in the ULCM design, we observed that the empirical allocation of patients in scenarios 1 to 6 in Table 4.1 still showed a pattern of equal randomization. This was because the control treatment served as the benchmark comparator for both treatments 1 and 2 throughout the trial, and early stopping rules were applied for treatments 1 and 2. Therefore, even if the response adaptive randomization scheme favored the investigational treatments with better survival outcomes, those treatment arms may be prematurely discontinued due to the early stopping rules. Conversely, the control treatment was always available for new patients unless the entire trial was prematurely terminated, which would bring the overall randomization ratios back to equal.

To further explore this matter, Table 4.2 presented the operating characteristics (OCs) of the ULCM, Conv 1, and Conv 2 designs without early stopping rules for futilities or superiority, achieved by setting $\zeta_u = 1$ and $\zeta_l = 0$. As anticipated, in the absence of early stopping rules, the ULCM design allocated more patients to treatments with higher progression-free survival (PFS) rates, thanks to the use of response adaptive randomization. Additionally, the ULCM design remained the superior design in terms of selecting the correct treatment. However, there was a tradeoff: the mean sample size reached the maximum value of 500 for all scenarios considered.

We performed additional sensitivity analyses to evaluate the ULCM design's robustness. For the main simulation studies in Table 4.1, we split the patients in stage I into two equal halves, with one half assigned to equal randomization and the other half to response adaptive randomization. We also tested other between-stage patients

allocation ratios, such as 25% and 75%, and summarized the correct treatment selection percentages in Figure 4.2. The results indicated that using 50% or 75% of patients in stage I leads to slightly higher percentages of correct treatment selection than using only 25% of patients in stage I. Furthermore, we generated PFS outcomes using the piecewise exponential distribution, as presented in Table 4.1. We also tested other data generating distributions such as the exponential and gamma distributions. Figure 4.3 summarized the correct treatment selection percentages obtained with different data generating distributions. The ULCM design was found to be robust with different underlying survival distributions. Additional sensitivity analysis results, including patients' allocation and mean sample size, were available in the Appendix C (Tables 6.3-6.4).

Table 4.1: The operating characteristics of the ULCM, Conv 1 and Conv 2 designs including percentages of treatment selection, patients' allocation and mean sample sizes. The results are based on 5,000 simulated trials.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 1				*Scenario 2*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
t=2	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
t=2	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
ULCM								
Treatment selection (%)								
Both treatments are effective	**96.6**	**97.0**	**97.2**	**97.2**	3.0	2.6	2.6	2.4
Only treatment 1 is effective	1.4	1.4	1.2	1.2	6.8	7.6	7.0	7.8
Only treatment 2 is effective	1.8	1.4	1.4	1.4	7.2	8.0	7.6	7.8
Neither of treatment is effective	0.2	0.2	0.2	0.2	**83.0**	**81.8**	**82.8**	**82.0**
Patient allocation (n)	**Total N = 361**				**Total N = 484**			
Control	36.2	24.0	35.9	24.0	50.7	33.7	50.6	34.2
Treatment 1	36.7	23.9	36.5	24.2	46.6	31.3	46.9	30.9
Treatment 2	35.9	23.7	35.6	23.9	47.8	31.7	47.4	32.2
Conv 1								
Treatment selection (%)								
Both treatments are effective	**87.2**	**80.6**	**86.0**	**81.2**	2.6	2.2	3.8	4.8
Only treatment 1 is effective	4.4	9.4	6.6	8.6	9.8	14.8	9.0	10.4
Only treatment 2 is effective	6.4	7.0	5.8	6.6	9.4	14.0	6.6	13.8
Neither of treatment is effective	2.0	3.0	1.6	3.6	**78.2**	**69.0**	**80.6**	**71.0**
Patient allocation (n)	**Total N = 420**				**Total N = 493**			
Control	41.8	27.8	41.4	27.8	51.4	34.1	50.7	32.9
Treatment 1	42.6	28.1	41.5	28.2	48.0	32.7	48.6	32.4
Treatment 2	41.5	28.5	41.5	29.2	49.5	32.3	48.5	32.2
Conv 2								
Treatment selection (%)								
Both treatments are effective	**76.4**	**63.8**	**70.0**	**64.8**	2.4	2.4	1.8	2.0
Only treatment 1 is effective	10.4	14.8	13.2	15.4	10.0	10.4	7.6	7.8
Only treatment 2 is effective	8.6	14.2	11.8	11.8	7.8	11.4	12.6	9.8
Neither of treatment is effective	4.6	7.2	5.0	8.0	**79.8**	**75.8**	**78.0**	**80.4**
Patient allocation (n)	**Total N = 449**				**Total N = 492**			
Control	44.1	29.8	43.9	29.8	51.7	34.6	51.9	34.2
Treatment 1	43.6	31.1	44.2	30.6	47.7	32.1	47.7	31.7
Treatment 2	44.4	31.3	44.5	31.9	48.0	32.3	48.4	32.2

Table 4.1 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 3				*Scenario 4*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.15	0.15	0.15	0.15
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.40	0.40	0.40	0.40
t=2	0.40	0.40	0.60	0.60	0.40	0.40	0.60	0.60
ULCM								
Treatment selection (%)								
Both treatments are effective	3.0	12.4	12.2	**93.8**	3.2	3.0	9.6	9.6
Only treatment 1 is effective	10.0	**83.2**	0.2	1.8	6.0	6.6	0.0	0.0
Only treatment 2 is effective	8.8	0.2	**85.0**	3.4	9.4	10.2	**88.0**	**88.0**
Neither of treatment is effective	**78.2**	4.2	2.6	1.0	**81.4**	**80.2**	2.4	2.4
Patient allocation (n)	**Total N = 467**				**Total N = 483**			
Control	50.2	35.4	53.8	25.3	51.2	33.7	53.4	35.7
Treatment 1	48.3	26.1	54.4	26.2	46.8	31.0	51.0	34.4
Treatment 2	48.7	36.2	37.4	24.7	48.4	32.7	39.0	25.9
Conv 1								
Treatment selection (%)								
Both treatments are effective	6.0	17.6	13.2	**86.0**	4.2	5.8	13.6	16.4
Only treatment 1 is effective	9.0	**72.4**	0.2	5.8	10.2	12.8	0.6	1.4
Only treatment 2 is effective	8.8	0.8	**81.8**	5.4	13.0	12.4	**78.6**	**72.6**
Neither of treatment is effective	**76.2**	9.2	4.8	2.8	**72.6**	**69.0**	7.2	9.6
Patient allocation (n)	**Total N = 473**				**Total N = 486**			
Control	50.5	34.3	52.4	27.4	51.2	33.4	52.2	33.7
Treatment 1	47.2	27.8	52.1	28.4	48.3	32.4	51.8	33.8
Treatment 2	49.4	34.6	40.7	28.2	48.3	32.2	40.5	28.2
Conv 2								
Treatment selection (%)								
Both treatments are effective	2.2	13.2	13.8	**63.8**	1.4	3.4	11.0	12.4
Only treatment 1 is effective	12.2	**63.0**	1.4	15.6	10.0	12.6	1.0	2.8
Only treatment 2 is effective	10.2	0.8	**70.2**	13.0	11.6	9.8	**73.0**	**62.4**
Neither of treatment is effective	**75.4**	23.0	14.6	7.6	**77.0**	**74.2**	15.0	22.4
Patient allocation (n)	**Total N = 483**				**Total N = 492**			
Control	51.1	34.2	51.4	30.2	51.2	33.9	51.8	33.8
Treatment 1	48.2	29.5	49.8	30.6	48.4	32.3	50.5	33.2
Treatment 2	48.7	34.0	44.5	30.9	48.6	32.6	44.3	31.0

Table 4.1 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 5				*Scenario 6*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.6	0.6	0.6	0.6
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.60	0.60	0.60	0.60
t=2	0.40	0.40	0.68	0.68	0.40	0.40	0.60	0.60
ULCM								
Treatment selection (%)								
Both treatments are effective	3.8	14.8	13.4	**96.8**	12.0	13.6	**94.8**	**94.8**
Only treatment 1 is effective	9.0	**82.0**	0.0	0.0	**86.0**	**84.4**	3.2	3.2
Only treatment 2 is effective	10.0	0.0	**86.6**	3.2	0.0	0.2	1.6	1.6
Neither of treatment is effective	**77.2**	3.2	0.0	0.0	2.0	1.8	0.4	0.4
Patient allocation (n)	**Total N = 455**				**Total N = 428**			
Control	51.2	35.8	56.7	23.5	53.1	35.0	38.2	25.6
Treatment 1	48.4	24.7	56.7	27.0	37.3	25.0	37.0	25.0
Treatment 2	47.7	35.2	28.4	19.4	52.9	35.0	38.1	25.5
Conv 1								
Treatment selection (%)								
Both treatments are effective	3.8	18.8	14.8	**89.0**	13.8	15.8	**87.0**	**80.6**
Only treatment 1 is effective	12.2	**70.0**	0.0	0.4	**77.6**	**72.4**	6.0	8.0
Only treatment 2 is effective	11.0	0.6	**85.2**	10.4	0.0	1.2	6.4	8.8
Neither of treatment is effective	**73.0**	10.6	0.0	0.2	8.6	10.6	0.6	2.6
Patient allocation (n)	**Total N = 461**				**Total N = 451**			
Control	50.8	34.3	54.6	26.3	51.5	34.1	41.0	27.6
Treatment 1	48.2	27.6	55.6	29.6	42.3	28.7	42.4	29.2
Treatment 2	48.3	34.1	30.1	21.3	51.3	34.2	40.4	28.4
Conv 2								
Treatment selection (%)								
Both treatments are effective	2.6	12.0	15.4	**77.4**	13.6	13.4	**76.4**	**60.2**
Only treatment 1 is effective	10.4	**62.8**	0.0	3.0	**71.2**	**65.6**	9.2	16.4
Only treatment 2 is effective	10.0	3.4	**81.4**	16.6	1.0	2.0	8.8	14.4
Neither of treatment is effective	**77.0**	21.8	3.2	3.0	14.2	19.0	5.6	9.0
Patient allocation (n)	**Total N = 474**				**Total N = 466**			
Control	50.9	33.9	53.2	28.4	51.5	33.2	43.8	30.4
Treatment 1	47.6	30.2	53.6	31.3	43.7	31.2	43.4	30.9
Treatment 2	48.7	34.0	36.0	26.3	49.7	33.1	45.0	30.7

Table 4.2: The operating characteristics of the ULCM, Conv 1 and Conv 2 designs without the early stopping rules.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 1				*Scenario 2*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
t=2	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
t=2	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
ULCM								
Treatment selection (%)								
Both treatments are effective	**93.8**	**94.0**	**93.8**	**93.8**	2.0	2.0	2.2	2.2
Only treatment 1 is effective	3.0	2.8	3.0	3.0	4.2	4.4	4.0	4.2
Only treatment 2 is effective	3.0	3.0	3.0	3.0	6.6	6.8	7.0	7.0
Neither of treatment is effective	0.2	0.2	0.2	0.2	**87.2**	**86.8**	**86.8**	**86.6**
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	40.6	26.9	40.8	27.0	50.9	33.4	51.1	33.4
Treatment 1	55.1	36.0	54.9	36.1	49.3	32.9	49.0	32.5
Treatment 2	55.3	36.3	54.8	36.4	49.9	33.9	50.2	33.3
Conv 1								
Treatment selection (%)								
Both treatments are effective	**83.4**	**78.4**	**83.8**	**74.0**	3.6	3.4	2.8	2.6
Only treatment 1 is effective	7.2	7.0	6.6	10.2	7.8	7.2	6.2	7.6
Only treatment 2 is effective	7.8	11.4	7.4	12.2	6.8	9.2	9.2	11.2
Neither of treatment is effective	1.6	3.2	2.2	3.6	**81.8**	**80.2**	**81.8**	**78.6**
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	40.2	26.0	39.7	25.8	49.1	31.9	49.3	31.5
Treatment 1	54.7	36.8	54.5	36.1	50.5	34.1	49.8	33.4
Treatment 2	55.7	37.3	55.6	37.4	50.6	34.2	51.0	34.5
Conv 2								
Treatment selection (%)								
Both treatments are effective	**70.2**	**55.4**	**66.4**	**56.8**	2.4	1.8	1.8	1.8
Only treatment 1 is effective	10.0	17.8	13.8	16.8	8.4	11.2	6.4	11.6
Only treatment 2 is effective	13.6	17.8	13.8	17.0	7.0	9.6	8.2	10.2
Neither of treatment is effective	6.2	9.0	6.0	9.4	**82.2**	**77.4**	**83.6**	**76.4**
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	41.6	27.4	41.6	27.1	49.9	32.9	50.4	33.0
Treatment 1	54.3	35.8	54.8	36.0	49.7	33.7	50.2	33.4
Treatment 2	54.7	36.3	53.5	36.8	50.3	33.8	49.8	32.9

Table 4.2 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 3				*Scenario 4*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.15	0.15	0.15	0.15
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.40	0.40	0.40	0.40
t=2	0.40	0.40	0.60	0.60	0.40	0.40	0.60	0.60
ULCM								
Treatment selection (%)								
Both treatments are effective	3.4	13.6	9.0	**92.8**	3.2	3.2	12.4	12.6
Only treatment 1 is effective	5.2	**82.0**	0.0	2.8	8.8	9.4	0.0	0.0
Only treatment 2 is effective	9.2	0.6	**87.4**	3.6	8.0	8.8	**84.4**	**84.2**
Neither of treatment is effective	**82.2**	3.8	3.6	0.8	**80.0**	**78.6**	3.2	3.2
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	49.9	28.7	43.4	26.6	49.6	32.7	43.8	29.5
Treatment 1	50.3	41.5	44.4	36.9	49.6	33.0	43.5	29.0
Treatment 2	49.5	29.5	62.3	36.9	51.0	33.9	63.0	41.6
Conv 1								
Treatment selection (%)								
Both treatments are effective	4.4	13.4	13.4	**79.6**	3.0	6.0	11.8	13.6
Only treatment 1 is effective	8.0	**78.0**	0.4	8.6	7.8	8.2	0.0	0.0
Only treatment 2 is effective	9.4	0.8	**81.0**	10.2	9.2	9.8	**82.6**	**77.8**
Neither of treatment is effective	**78.2**	7.8	5.2	1.6	**80.0**	**76.0**	5.6	8.6
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	48.8	28.1	42.8	26.3	49.7	31.3	43.8	27.9
Treatment 1	51.1	40.8	44.7	36.5	49.8	34.3	45.2	30.1
Treatment 2	50.0	30.9	62.7	37.3	50.9	34.4	61.9	40.7
Conv 2								
Treatment selection (%)								
Both treatments are effective	1.6	12.4	12.4	**55.2**	3.2	2.8	10.8	12.4
Only treatment 1 is effective	9.4	**61.6**	1.0	19.2	8.8	11.0	1.4	2.6
Only treatment 2 is effective	8.0	3.2	**69.6**	16.4	8.8	9.0	**72.8**	**61.4**
Neither of treatment is effective	**81.0**	22.8	17.0	9.2	**79.2**	**77.2**	15.0	23.6
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	49.9	29.6	44.6	27.6	49.4	33.0	45.1	29.8
Treatment 1	50.0	40.0	45.1	36.6	49.8	33.6	45.2	30.5
Treatment 2	49.7	30.3	60.7	35.9	50.5	33.5	60.2	39.5

Table 4.2 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 5				*Scenario 6*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.6	0.6	0.6	0.6
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.60	0.60	0.60	0.60
t=2	0.40	0.40	0.68	0.68	0.40	0.40	0.60	0.60
ULCM								
Treatment selection (%)								
Both treatments are effective	2.6	10.0	10.8	**95.2**	11.8	11.4	**92.4**	**92.8**
Only treatment 1 is effective	7.8	**85.2**	0.0	0.0	**85.0**	**85.8**	4.4	4.4
Only treatment 2 is effective	7.0	0.0	**89.2**	4.8	0.0	0.0	2.4	2.0
Neither of treatment is effective	**82.6**	4.8	0.0	0.0	3.2	2.8	0.8	0.8
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	50.3	28.4	37.9	23.9	44.1	28.7	40.7	27.1
Treatment 1	49.7	41.7	38.9	32.2	61.5	40.2	53.7	35.8
Treatment 2	49.9	29.4	73.7	44.0	45.4	30.4	55.1	37.4
Conv 1								
Treatment selection (%)								
Both treatments are effective	3.0	12.6	14.6	**82.0**	13.2	14.0	**82.2**	**75.8**
Only treatment 1 is effective	10.6	**74.0**	0.0	0.8	**77.0**	**71.6**	7.2	10.4
Only treatment 2 is effective	8.4	1.0	**84.6**	16.8	0.6	0.4	8.0	10.6
Neither of treatment is effective	**78.0**	12.4	0.8	0.4	9.2	14.0	2.6	3.2
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	49.4	28.4	37.3	22.7	43.8	28.3	39.4	25.9
Treatment 1	51.2	41.4	38.8	31.2	61.4	40.5	54.2	36.3
Treatment 2	50.5	30.3	73.3	45.5	45.5	31.1	56.1	37.5
Conv 2								
Treatment selection (%)								
Both treatments are effective	1.8	13.2	11.6	**68.2**	11.6	10.2	**70.4**	**55.0**
Only treatment 1 is effective	9.2	**64.6**	0.6	7.0	**70.6**	**65.0**	13.0	16.8
Only treatment 2 is effective	8.4	1.2	**83.4**	21.8	1.0	1.6	12.6	18.4
Neither of treatment is effective	**80.6**	21.0	4.4	3.0	16.8	23.2	4.0	9.8
Patient allocation (n)	**Total N = 500**				**Total N = 500**			
Control	49.8	29.4	42.1	25.9	44.8	29.4	41.3	27.6
Treatment 1	49.9	40.5	42.9	34.0	59.9	39.4	54.2	36.4
Treatment 2	50.1	30.0	65.2	40.1	45.2	30.3	54.8	36.7

Figure 4.2: The percentages of correct treatment selection within each subgroup of ULCM design with different ratios of number of patients between Stage I and Stage II.

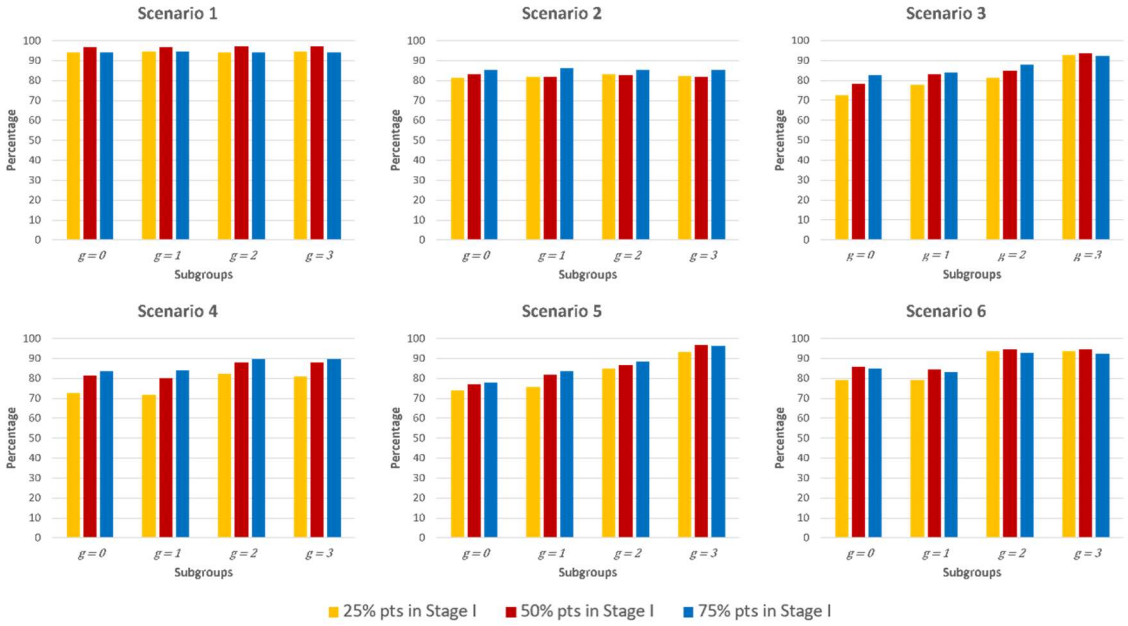
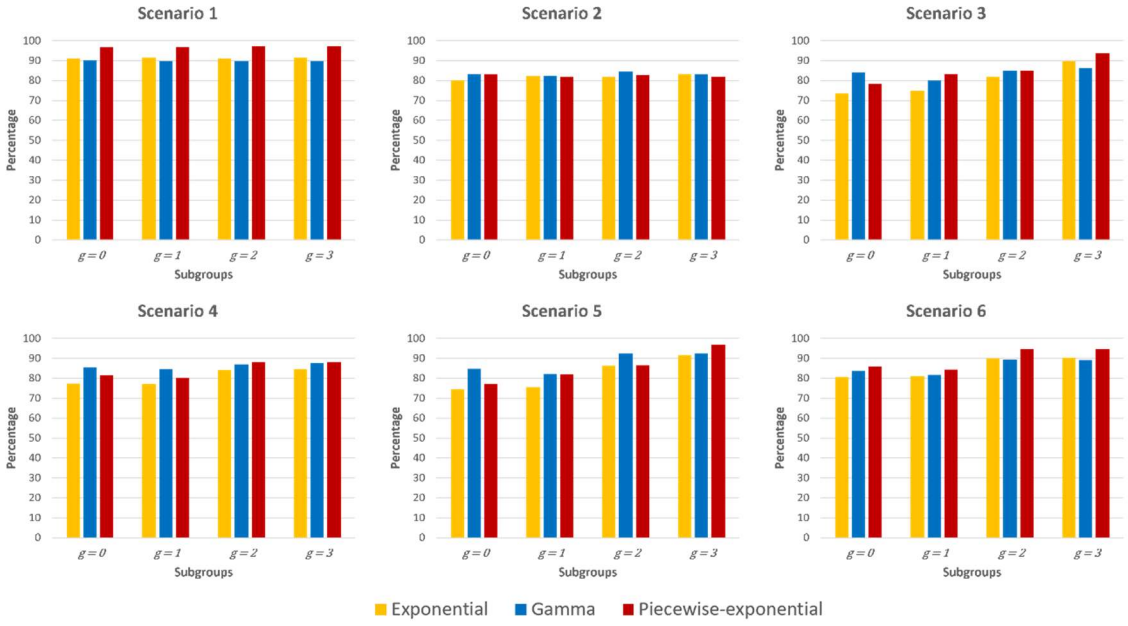


Figure 4.3: The percentages of correct treatment selection within each subgroup of ULCM design with different data generating distributions of PFS.



4.4 Conclusion

In this paper, we introduce the Bayesian response-adaptive umbrella trial design (ULCM) for evaluating subgroup-specific treatment effects. Our design employs a mixture regression model that jointly models short-term and long-term response outcomes to facilitate treatment effect estimation. We also propose a latent class model that groups biomarker-based subgroups into underlying classes for information sharing. Our simulation results demonstrate the ULCM design's desirable performance.

The ULCM is a modular design that can be customized to fit specific clinical settings. For instance, in situations where multiple targeted agents are designed for the same biomarker, a hierarchical model can be employed to leverage information from these agents for agent-specific biomarker-based subgroups. Furthermore, if the biomarker expression is a continuous outcome instead of binary, we can use an additional latent classification model equipped with underlying cutoff values determined by the data.

Although related, a different problem is the basket trial, which evaluates a single investigational drug across various patient populations defined by characteristics such as disease stage, histology, or biomarkers (Woodcock and LaVange, 2017). We are keen to extend our proposed Bayesian latent class model to the basket trial design. R code for implementing the ULCM design is available at https://github.com/yongzang2020.

Chapter 5

Conclusion and Discussion

Delayed outcomes are common and critical issues in early phase dose finding trials. This dissertation proposed two novel phase I dose-finding designs that can handle delayed outcomes while maintaining the simplicity and transparency of rule-based designs. At the same time, they are constructed on a statistically rigorous Bayesian probability model, enabling it to incorporate the information from the pending patients' remaining follow-up time.

In Chapter 2, we proposed a new phase I dosing finding trial design, T-3+3, to handle delayed outcomes in the 3+3 design. Like most algorithm-based designs, T-3+3 is easy to implement because its decision rule can be pre-tabulated, which dramatically facilitates the trial conduct without intensive statistical computation. We showed that the T-3+3 design is more efficient in MTD identification and patient allocation than the rolling six design. In the meantime, it further shortens the trial duration. Compared to the conventional 3+3 design, the T-3+3 design supports continuous accrual while maintaining patient safety and accuracy in identifying the MTD, making it a practical phase I clinical trial design to expedite early-phase drug development. Simulations confirmed that the T-3+3 design was robust and yielded desirable operating characteristics under various settings in terms of accrual rate, percentage of toxicity occurring in the latter half of the assessment window, and time-to-toxicity distribution.

In Chapter 3, we aimed to assess the suitability of our Bayesian probability model, which effectively handles delayed outcomes, for a more sophisticated phase I trial design. We proposed the T-i3+3 design to accommodate delayed outcomes in the i3+3 design. The T-i3+3 design shares desirable features of simplicity and transparency with the i3+3 design, including a pre-tabulated decision rule. One of the most significant advantages of the T-i3+3 design compared to the i3+3 design is that it does not require frequent suspension of the trial in the presence of delayed outcomes, resulting in a significant reduction in the trial duration.

Our approach of using Bayesian probability model to handle delayed outcomes is approved to be applicable to any algorithm-based designs as they share the same framework of performing dose escalation/de-escalation based on pre-determined observation of DLTs. We look forward to proposing more models using this approach to address many of the unmet challenges that are present in early dose finding designs.

Our next proposed trial design aims to enhance the performance of master protocol trials for personalized medicine. Due to advances in understanding of cancer biology, targeted therapies are being developed to offer patients more effective treatment and better outcomes using personalized predictive biomarkers. The umbrella trial is a clinical trial strategy that accommodates the paradigm shift towards personalized medicine by evaluating multiple investigational drugs in different subgroups of patients with the same disease (Woodcock and LaVange, 2017; Dean *et al*., 2020). These subgroups are defined by various biomarkers, such as genetic or molecular aberrations.

In chapter 4, we proposed the Bayesian response-adaptive umbrella trial design (ULCM) to address the challenges posed by extended follow-up time required for PFS and the reduction in statistical power caused by small sample sizes in subgroup-specific treatment groups in oncology umbrella trial design. For the first challenge, we used the short-term CBR as a surrogate marker for PFS and developed a mixture regression model with parametric Weibull distribution for the baseline hazard to borrow information between the CBR and PFS. To overcome subgroup-specific sample size limitations, we also proposed a data-driven Bayesian latent class model. This model adaptively groups subgroups into latent classes based on overall data heterogeneities, and then pools response outcomes within the latent classes. Our simulation results demonstrate that the ULCM design exhibits desirable performance and is found to be robust to different underlying survival distributions.

Relying on the assumption of a relationship between PFS and CBR may be a limitation, as it may not always hold true in practice. Therefore, it is important to obtain input from clinicians regarding the model justification when implementing the proposed method to design new trials.

The ULCM offers a modular design that can be easily tailored to suit specific clinical settings. For example, in scenarios where certain treatments are not available to a particular biomarker-based subgroup, we can employ a modified latent class model exclusively for that subgroup, while keeping the other subgroups unaffected. Moreover, In the case where the biomarker expression yields a continuous outcome rather than a

binary one, we can employ an additional latent classification model that incorporates data-derived cutoff values.

In addition, a related but different trial design is the basket trial, which evaluates a single investigational drug across various patient populations defined by characteristics such as disease stage, histology, or biomarkers (Woodcock and LaVange, 2017). We are keen to extend our proposed Bayesian latent class model to the basket trial design, which could provide a flexible and efficient approach to treatment selection in this setting.

Chapter 6

Appendices

6.1 Appendix A (Chapter 2)

Derivation of the beta-binomial distribution

Let p be the probability of developing the DLT for any patient treated at the current dose. Then, using the weighted likelihood function method (Cheung and Chappell, 2000) and the Taylor expansion approximation (Lin and Yuan, 2020), the likelihood function can be approximated as

$$L \approx p^{r_1}(1-p)^{n_1-r_1}\prod_{i=1}^{n_2}\left(1-\frac{v_i}{U}p\right)$$

$$\approx p^{r_1}(1-p)^{n_1-r_1}\sum_{i=1}^{n_2}(1-p)^{\frac{v_i}{U}}$$

$$= p^{r_1}(1-p)^{n_1-r_1+n_2\frac{\bar{v}}{U}}$$

After assigning p a non-informative uniform prior distribution $U(0,1)$, it is immediately clear that p follows a beta posterior distribution as

$$p \sim beta(r_1+1, n_1-r_1+n_2\frac{\bar{v}}{U}+1),$$

with the density function $\pi(p|\mathcal{D}) = \frac{p^{r_1}(1-p)^{n_1-r_1+n_2\frac{\bar{v}}{U}}}{\mathrm{B}(r_1+1, n_1-r_1+n_2\frac{\bar{v}}{U}+1)}$.

When p is fixed, r_2 follows a binomial distribution with the probability $\pi(k|p) = Pr(r_2 = k|p)$ expressed as

$$\pi(k|p) = \frac{n_2!}{k!\,(n_2 - k)!} p^k (1-p)^{n_2 - k}$$

Then, by integrating out $\pi(p|\mathcal{D})$, the posterior distribution of r_2, referred to $\pi(k|\mathcal{D})$ can be expressed as

$$\pi(k|\mathcal{D}) = \int_0^1 \pi(k|p)\pi(p|\mathcal{D})dp$$

$$= \frac{n_2!}{k!\,(n_2 - k)!} \frac{\mathrm{B}(k + r_1 + 1, n_2 - k + n_1 - r_1 + n_2 \frac{\overline{v}}{U} + 1)}{\mathrm{B}(r_1 + 1, n_1 - r_1 + n_2 \frac{\overline{v}}{U} + 1)}$$

Figure 6.1: The patient allocation percentages at the MTD of the 3+3, Rolling 6 and T-3+3 designs with different patient enrollment rates.

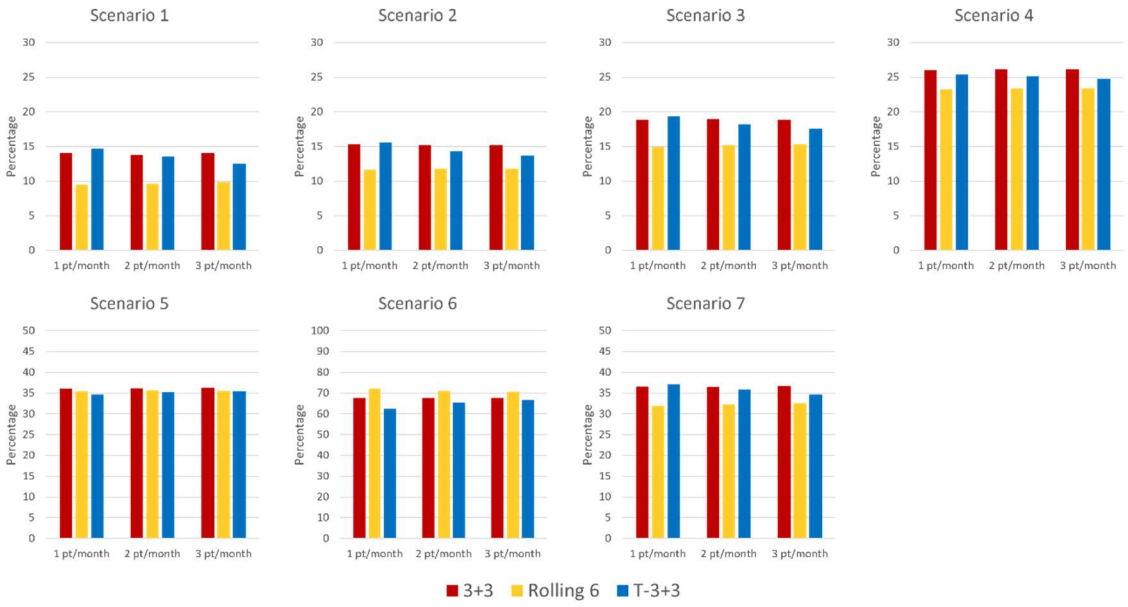


Figure 6.2: The patient allocation percentages at the MTD of the 3+3, Rolling 6 and T-3+3 designs with different event happening probabilities in the latter half of the assessment window.

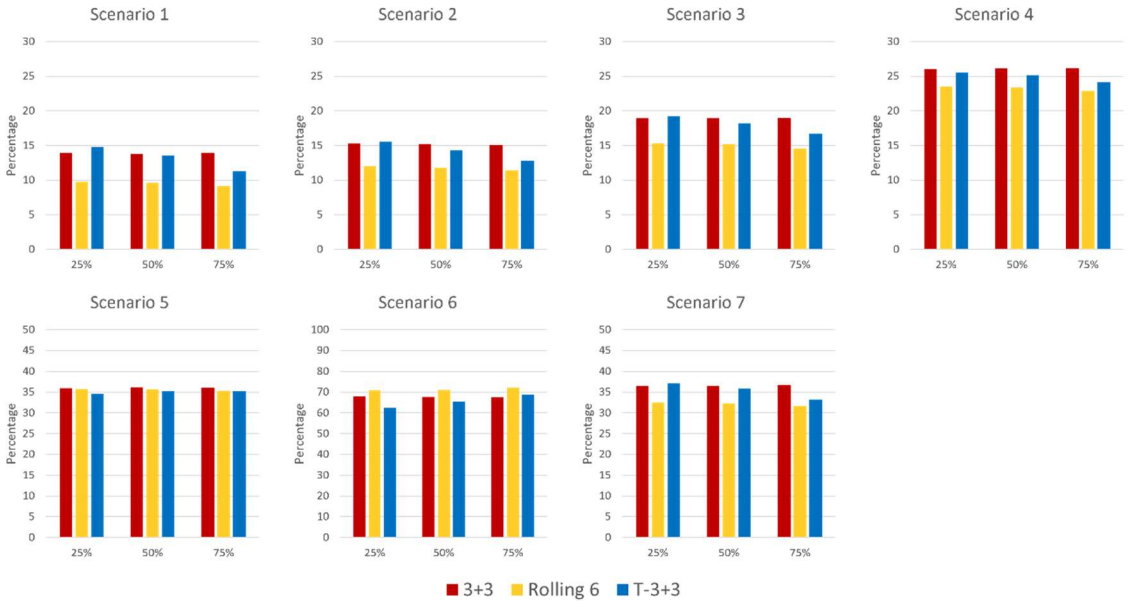
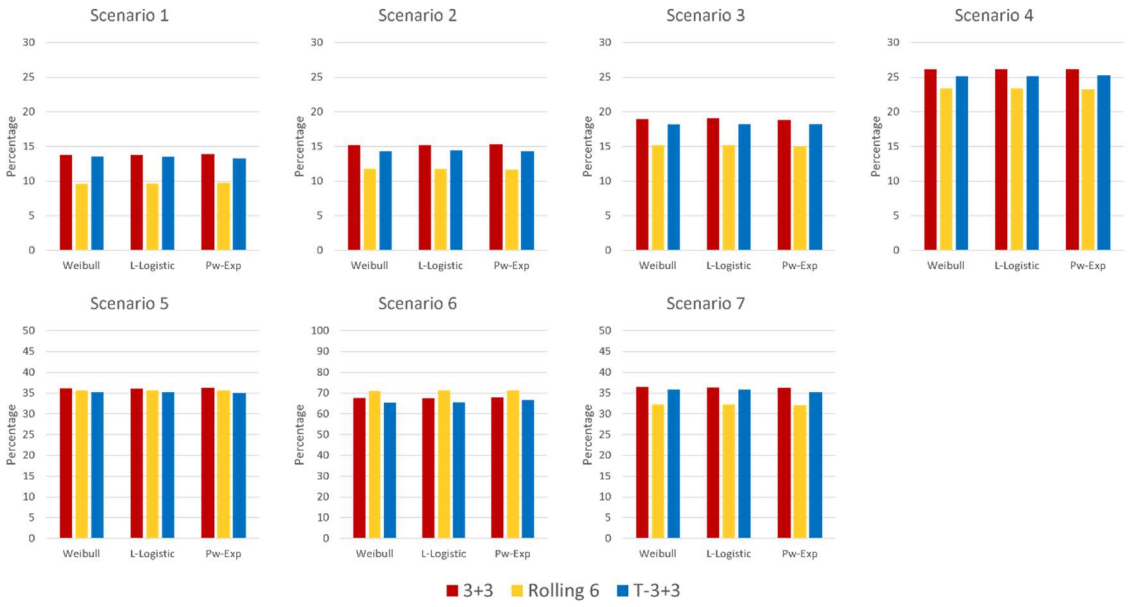


Figure 6.3: The patient allocation percentages at the MTD of the 3+3, Rolling 6 and T-3+3 designs with different time-to-event distributions.



6.2 Appendix B (Chapter 3)

Table 6.1: The operating characteristics of the i3+3 and T-i3+3 designs, including percentages of MTD selection (Selection (%)) and patients' allocation (Patients (%)) at each dose level, number of patients in the trial and trial duration, based on 10,000 simulated trials. The targeted DLT rate is 0.1.

Designs		Dose level							Duration (month)	No. of Patients	No. of DLTs
		0	1	2	3	4	5	6			
					Scenario 1						
	True toxicity		0.01	0.02	0.03	0.04	0.06	**0.10**			
i3 + 3	Selection(%)	0.1	0.3	1.4	4.4	9.1	22.9	**61.9**	53.2	36.0	2.1
	Patients (%)		9.7	11.2	12.6	13.9	17.3	35.3			
T-i3 + 3	Selection(%)	0.0	0.2	1.4	3.6	7.5	23.4	**63.9**	31.1	36.0	2.2
	Patients (%)		9.6	10.8	12.0	12.8	16.5	38.4			
					Scenario 2						
	True toxicity		0.02	0.04	0.06	0.08	**0.10**	0.28			
i3 + 3	Selection(%)	0.1	1.8	7.0	14.8	21.8	**46.6**	7.9	52.9	36.0	3.4
	Patients (%)		12.2	15.7	17.7	17.6	23.0	13.9			
T-i3 + 3	Selection(%)	0.2	1.4	5.5	12.2	20.4	**51.2**	9.2	30.7	36.0	3.7
	Patients (%)		11.5	14.2	15.9	16.4	24.0	17.9			
					Scenario 3						
	True toxicity		0.01	0.03	0.05	**0.10**	0.20	0.34			
i3 + 3	Selection(%)	0.0	0.6	4.8	21.5	**46.4**	24.4	2.3	52.9	36.0	3.9
	Patients (%)		10.2	13.9	20.2	27.4	20.6	7.8			
T-i3 + 3	Selection(%)	0.0	0.4	3.8	18.9	**46.8**	27.7	2.4	30.5	36.0	4.4
	Patients (%)		9.9	12.6	17.9	25.7	22.9	10.9			
					Scenario 4						
	True toxicity		0.02	0.05	**0.09**	0.14	0.16	0.20			
i3 + 3	Selection(%)	0.2	2.7	15.4	**30.5**	24.0	17.0	10.2	52.9	35.9	3.7
	Patients (%)		13.1	20.6	24.6	19.3	12.8	9.6			
T-i3 + 3	Selection(%)	0.1	2.3	12.6	**26.5**	23.4	22.4	12.7	30.2	36.0	3.9
	Patients (%)		12.3	17.7	21.8	18.6	15.4	14.1			

Table 6.1 continues.

Designs		Dose level							Duration (month)	No. of Patients	No. of DLTs
		0	1	2	3	4	5	6			
					Scenario 5						
	True toxicity		0.05	**0.10**	0.15	0.20	0.25	0.30			
i3 + 3	Selection(%)	1.3	18.2	**32.4**	27.4	14.5	5.1	1.0	52.3	35.6	4.5
	Patients (%)		24.9	29.4	23.3	13.6	6.3	2.6			
T-i3 + 3	Selection(%)	1.2	15.9	**29.3**	27.7	16.8	7.6	1.6	29.2	35.7	4.9
	Patients (%)		21.8	26.1	23.0	15.4	8.9	4.8			
					Scenario 6						
	True toxicity		**0.11**	0.15	0.21	0.25	0.30	0.35			
i3 + 3	Selection(%)	12.4	**42.7**	27.2	11.8	4.4	1.3	0.2	48.6	33.2	5.1
	Patients (%)		44.9	29.1	15.6	6.9	2.6	0.9			
T-i3 + 3	Selection(%)	11.0	**37.0**	27.5	15.1	7.2	2.0	0.2	28.0	33.6	5.6
	Patients (%)		37.0	28.2	18.3	9.9	4.7	1.9			
					Scenario 7						
	True toxicity		0.02	0.03	0.04	0.05	0.06	**0.07**			
i3 + 3	Selection(%)	0.1	1.3	3.3	7.1	9.5	14.1	**64.7**	53.2	36.0	1.8
	Patients (%)		11.7	12.9	13.8	13.8	13.9	33.8			
T-i3 + 3	Selection(%)	0.1	1.0	2.6	5.6	7.7	13.6	**69.5**	30.9	36.0	1.9
	Patients (%)		11.2	12.1	12.7	12.9	13.4	37.7			
					Scenario 8						
	True toxicity		0.21	0.32	0.41	0.50	0.60	0.70			
i3 + 3	Selection(%)	61.8	36.3	1.9	0.1	0.0	0.0	0.0	34.0	23.5	5.7
	Patients (%)		76.5	18.6	4.2	0.7	0.1	0.0			
T-i3 + 3	Selection(%)	57.6	39.1	3.0	0.3	0.0	0.0	0.0	22.5	24.9	6.4
	Patients (%)		67.7	23.0	7.3	1.7	0.3	0.0			

Table 6.2: The operating characteristics of the i3+3 and T-i3+3 designs, including percentages of MTD selection (Selection (%)) and patients' allocation (Patients (%)) at each dose level, number of patients in the trial and trial duration, based on 10,000 simulated trials. The targeted DLT rate is 0.4.

Designs		Dose level							Duration (month)	No. of Patients	No. of DLTs
		0	1	2	3	4	5	6			
					Scenario 1						
	True toxicity		0.06	0.12	0.18	0.24	0.30	**0.40**			
i3 + 3	Selection(%)	0.0	0.2	0.7	2.2	7.7	27.7	**61.5**	51.8	36.0	10.0
	Patients (%)		8.6	9.4	11.1	13.6	19.5	37.7			
T-i3 + 3	Selection(%)	0.0	0.2	0.9	2.6	9.2	26.5	**60.6**	25.8	36.0	10.2
	Patients (%)		8.6	9.2	10.3	12.2	18.8	40.9			
					Scenario 2						
	True toxicity		0.05	0.10	0.20	0.30	**0.40**	0.56			
i3 + 3	Selection(%)	0.0	0.1	0.9	5.9	26.4	**50.6**	16.1	51.3	36.0	11.7
	Patients (%)		8.5	9.6	13.8	20.9	27.1	20.1			
T-i3 + 3	Selection(%)	0.0	0.1	1.3	6.9	27.2	**49.5**	15.1	26.3	36.0	12.3
	Patients (%)		8.5	9.5	12.2	17.5	26.5	25.8			
					Scenario 3						
	True toxicity		0.15	0.20	0.32	**0.42**	0.52	0.64			
i3 + 3	Selection(%)	0.4	0.9	6.6	28.7	**39.1**	21.6	2.8	50.8	35.9	13.3
	Patients (%)		10.2	15.2	23.7	25.3	17.9	7.8			
T-i3 + 3	Selection(%)	0.4	1.5	8.1	28.3	**37.8**	21.8	2.2	27.0	35.9	14.2
	Patients (%)		9.7	13.4	19.4	22.5	20.8	14.2			
					Scenario 4						
	True toxicity		0.14	0.30	**0.40**	0.58	0.72	0.86			
i3 + 3	Selection(%)	0.2	4.6	25.7	**54.2**	14.5	0.7	0.0	50.6	35.9	14.2
	Patients (%)		14.0	25.1	35.2	19.7	5.3	0.7			
T-i3 + 3	Selection(%)	0.3	5.1	28.0	**49.6**	16.2	0.8	0.0	26.9	35.9	15.8
	Patients (%)		12.8	20.5	29.7	22.0	10.3	4.7			

Table 6.2 continues.

Designs		Dose level 0	1	2	3	4	5	6	Duration (month)	No. of Patients	No. of DLTs
					Scenario 5						
	True toxicity		0.30	**0.40**	0.48	0.56	0.66	0.70			
i3 + 3	Selection(%)	3.7	22.9	**36.1**	26.0	9.3	1.8	0.2	49.0	34.9	14.8
	Patients (%)		29.0	30.6	23.0	12.0	4.3	1.1			
T-i3 + 3	Selection(%)	3.2	24.4	**34.6**	25.7	10.3	1.8	0.1	27.2	35.0	16.1
	Patients (%)		22.9	24.7	21.9	16.1	9.1	5.3			
					Scenario 6						
	True toxicity		**0.38**	0.46	0.55	0.64	0.68	0.72			
i3 + 3	Selection(%)	10.8	**42.2**	32.0	12.2	2.3	0.6	0.0	46.3	33.2	15.1
	Patients (%)		42.8	31.3	17.4	6.3	1.7	0.4			
T-i3 + 3	Selection(%)	8.1	**43.4**	31.9	13.0	3.0	0.6	0.1	26.6	34.0	16.7
	Patients (%)		32.9	26.9	19.6	11.1	6.1	3.4			
					Scenario 7						
	True toxicity		0.03	0.05	0.10	0.15	0.20	**0.32**			
i3 + 3	Selection(%)	0.0	0.0	0.1	0.3	1.2	9.5	**89.0**	52.3	36.0	7.9
	Patients (%)		8.4	8.5	8.9	9.8	14.4	50.0			
T-i3 + 3	Selection(%)	0.0	0.0	0.1	0.4	1.6	9.3	**88.5**	24.4	36.0	7.9
	Patients (%)		8.4	8.5	8.8	9.5	14.3	50.6			
					Scenario 8						
	True toxicity		0.54	0.65	0.75	0.85	0.90	0.95			
i3 + 3	Selection(%)	59.5	38.6	1.9	0.0	0.0	0.0	0.0	32.2	23.7	13.5
	Patients (%)		77.3	18.3	3.9	0.4	0.0	0.0			
T-i3 + 3	Selection(%)	51.2	46.1	2.6	0.1	0.0	0.0	0.0	22.2	26.7	16.1
	Patients (%)		61.6	23.2	9.7	3.6	1.3	0.5			

Figure 6.4: The trial duration in month of the i3+3 and T-i3+3 designs with different event happening probabilities in the latter half of the assessment window.

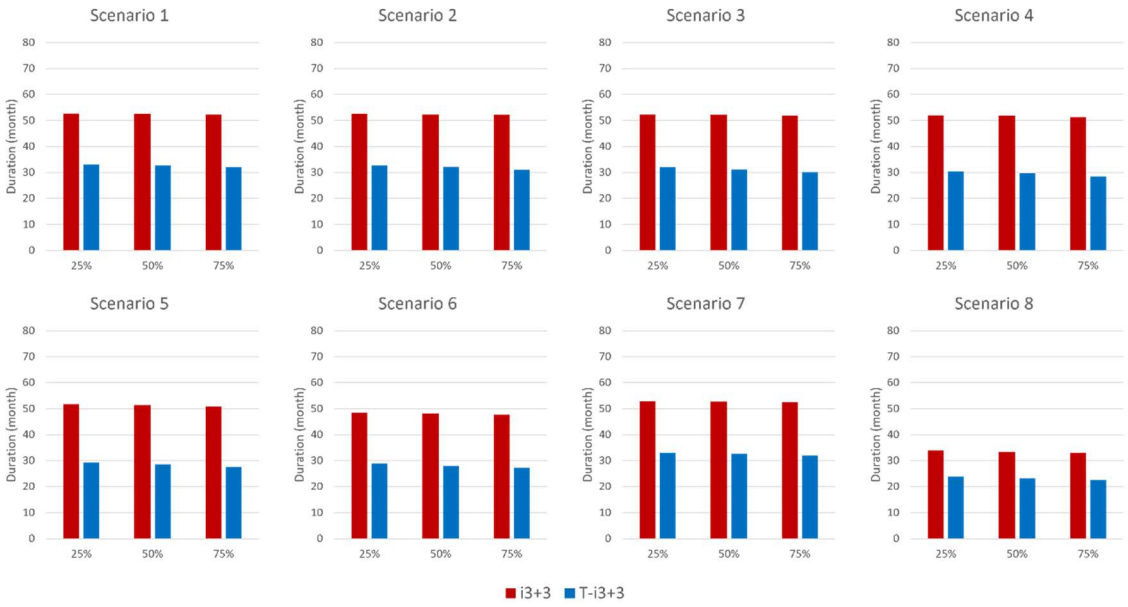


Figure 6.5: The correct MTD selection percentages of the i3+3 and T-i3+3 designs with different event happening probabilities in the latter half of the assessment window.

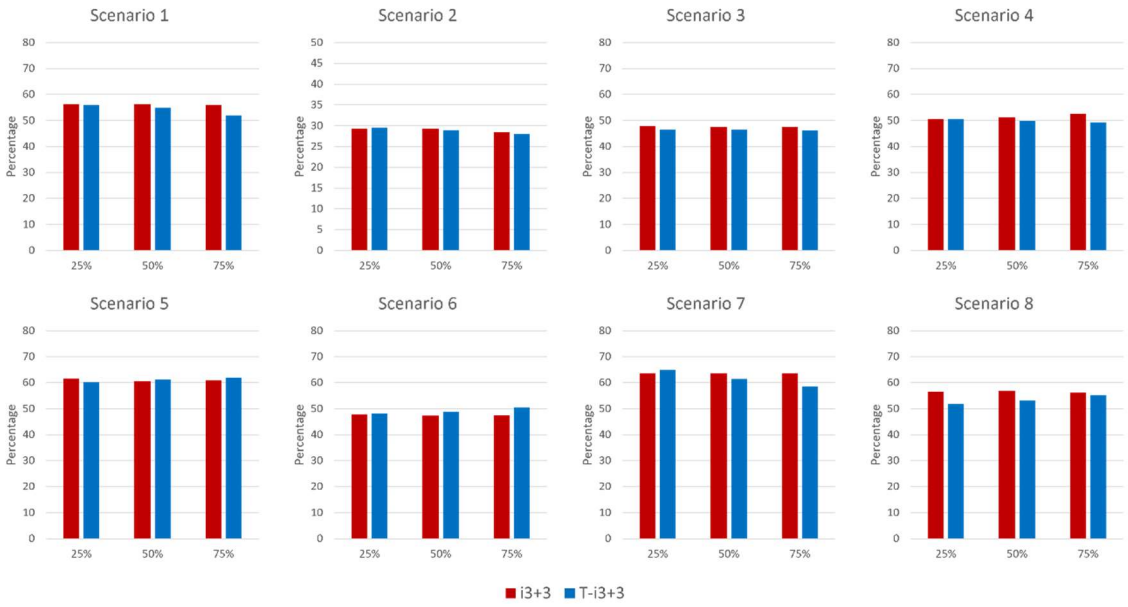


Figure 6.6: The trial duration in month of the i3+3 and T-i3+3 designs with different time-to-event distributions.

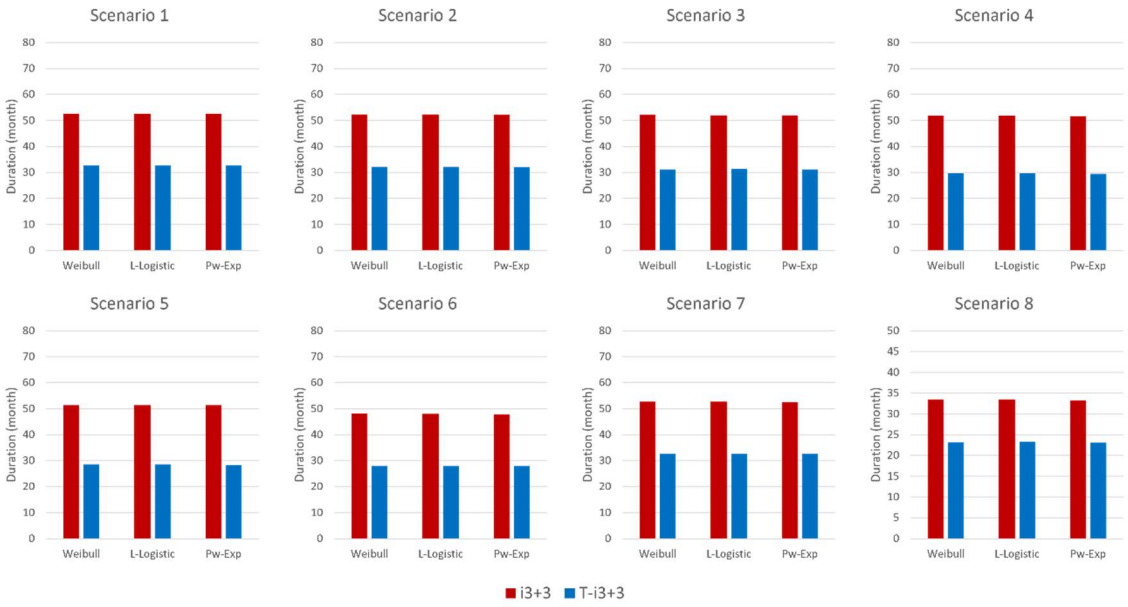
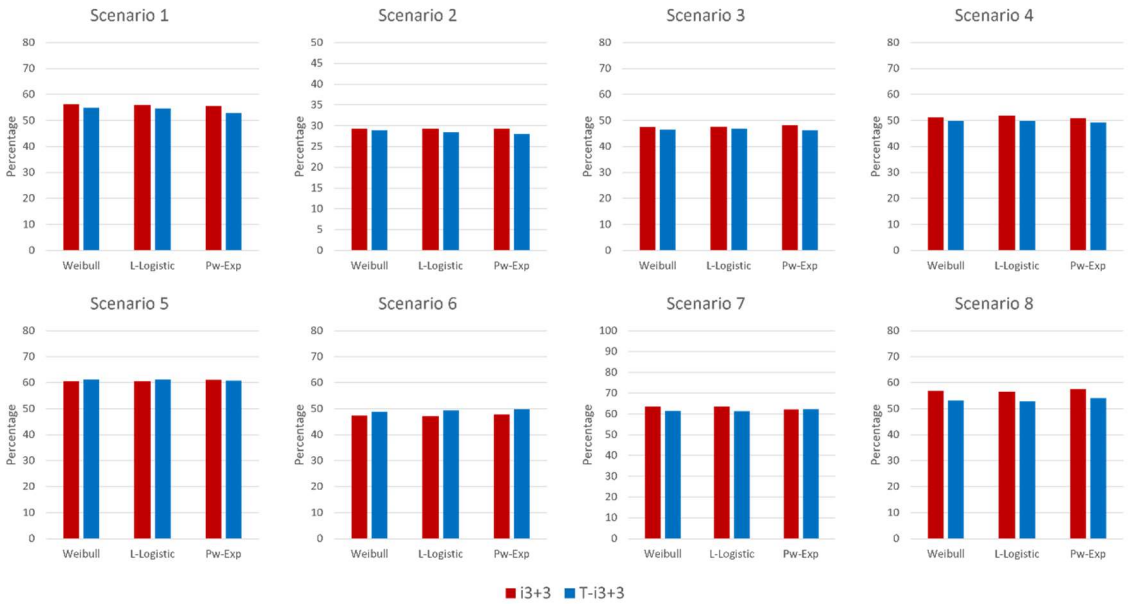


Figure 6.7: The correct MTD selection percentages of the i3+3 and T-i3+3 designs with different time-to-event distributions.



6.3 Appendix C (Chapter 4)

Table 6.3: The operating characteristics of the ULCM design including percentages of treatment selection, patients' allocation and mean sample sizes with different between-stages patients' allocation ratios.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 1				*Scenario 2*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
t=2	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
t=2	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
ULCM with 25% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	**94.2**	**94.8**	**94.0**	**94.4**	3.2	3.0	3.0	3.4
Only treatment 1 is effective	1.8	1.8	2.2	2.2	8.4	8.6	7.6	8.2
Only treatment 2 is effective	2.6	2.0	2.4	2.0	6.8	6.4	6.4	6.0
Neither of treatment is effective	1.4	1.4	1.4	1.4	**81.6**	**82.0**	**83.0**	**82.4**
Patient allocation (n)	**Total N = 333**				**Total N = 464**			
Control	32.3	21.6	31.8	21.4	50.7	34.1	51.0	33.2
Treatment 1	34.3	22.2	34.2	22.5	44.2	29.2	45.0	29.4
Treatment 2	34.3	22.8	33.3	22.4	44.4	30.0	43.6	28.8
ULCM with 50% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	**96.6**	**97.0**	**97.2**	**97.2**	3.0	2.6	2.6	2.4
Only treatment 1 is effective	1.4	1.4	1.2	1.2	6.8	7.6	7.0	7.8
Only treatment 2 is effective	1.8	1.4	1.4	1.4	7.2	8.0	7.6	7.8
Neither of treatment is effective	0.2	0.2	0.2	0.2	**83.0**	**81.8**	**82.8**	**82.0**
Patient allocation (n)	**Total N = 361**				**Total N = 484**			
Control	36.2	24.0	35.9	24.0	50.7	33.7	50.6	34.2
Treatment 1	36.7	23.9	36.5	24.2	46.6	31.3	46.9	30.9
Treatment 2	35.9	23.7	35.6	23.9	47.8	31.7	47.4	32.2
ULCM with 75% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	**94.2**	**94.4**	**94.0**	**94.2**	1.4	1.4	1.6	1.6
Only treatment 1 is effective	3.0	2.8	3.0	3.0	7.0	6.8	6.4	6.6
Only treatment 2 is effective	1.4	1.4	1.8	1.6	6.2	5.6	6.4	6.4
Neither of treatment is effective	1.4	1.4	1.2	1.2	**85.4**	**86.2**	**85.6**	**85.4**
Patient allocation (n)	**Total N = 429**				**Total N = 497**			
Control	43.4	28.5	43.3	28.1	50.9	33.5	51.4	34.0
Treatment 1	43.6	28.1	42.9	28.7	48.9	32.7	48.6	32.7
Treatment 2	42.8	28.6	42.7	28.5	49.3	33.1	48.9	33.1

Table 6.3 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 3				*Scenario 4*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.15	0.15	0.15	0.15
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.40	0.40	0.40	0.40
t=2	0.40	0.40	0.60	0.60	0.40	0.40	0.60	0.60
ULCM with 25% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	3.8	17.0	14.6	**92.6**	5.2	5.6	13.8	15.4
Only treatment 1 is effective	11.4	**78.0**	0.0	2.4	9.6	9.8	0.0	0.0
Only treatment 2 is effective	12.4	0.2	**81.2**	3.2	12.6	12.6	**82.4**	**80.8**
Neither of treatment is effective	**72.4**	4.8	4.2	1.8	**72.6**	**72.0**	3.8	3.8
Patient allocation (n)	**Total N = 445**				**Total N = 466**			
Control	50.8	34.6	51.6	23.2	50.9	33.9	53.0	34.6
Treatment 1	45.6	23.9	51.2	24.2	44.3	29.4	49.2	32.8
Treatment 2	46.0	34.0	36.4	23.9	46.3	32.0	35.9	23.6
ULCM with 50% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	3.0	12.4	12.2	**93.8**	3.2	3.0	9.6	9.6
Only treatment 1 is effective	10.0	**83.2**	0.2	1.8	6.0	6.6	0.0	0.0
Only treatment 2 is effective	8.8	0.2	**85.0**	3.4	9.4	10.2	**88.0**	**88.0**
Neither of treatment is effective	**78.2**	4.2	2.6	1.0	**81.4**	**80.2**	2.4	2.4
Patient allocation (n)	**Total N = 467**				**Total N = 483**			
Control	50.2	35.4	53.8	25.3	51.2	33.7	53.4	35.7
Treatment 1	48.3	26.1	54.4	26.2	46.8	31.0	51.0	34.4
Treatment 2	48.7	36.2	37.4	24.7	48.4	32.7	39.0	25.9
ULCM with 75% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	1.8	10.8	8.8	**92.4**	2.4	2.4	8.0	8.0
Only treatment 1 is effective	8.0	**84.0**	0.0	2.4	5.6	5.6	0.0	0.0
Only treatment 2 is effective	7.6	0.0	**87.8**	4.2	8.2	8.0	**89.6**	**89.6**
Neither of treatment is effective	**82.6**	5.2	3.4	1.0	**83.8**	**84.0**	2.4	2.4
Patient allocation (n)	**Total N = 483**				**Total N = 495**			
Control	50.7	34.8	52.6	29.2	51.0	33.5	53.1	35.2
Treatment 1	48.7	28.9	52.5	29.2	49.3	32.8	51.9	34.9
Treatment 2	49.4	35.1	43.3	29.0	49.3	32.8	43.0	28.6

Table 6.3 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 5				*Scenario 6*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.6	0.6	0.6	0.6
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.60	0.60	0.60	0.60
t=2	0.40	0.40	0.68	0.68	0.40	0.40	0.60	0.60
ULCM with 25% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	5.6	17.4	14.6	**93.2**	17.8	17.6	**93.8**	**93.8**
Only treatment 1 is effective	9.4	**75.8**	0.0	0.0	**79.4**	**79.4**	3.4	3.2
Only treatment 2 is effective	11.2	0.0	**85.0**	6.4	0.4	0.2	2.6	2.6
Neither of treatment is effective	**73.8**	6.8	0.4	0.4	2.4	2.8	0.2	0.4
Patient allocation (n)	**Total N = 427**				**Total N = 400**			
Control	51.4	34.6	52.8	20.4	52.1	34.2	33.1	21.9
Treatment 1	44.6	23.9	54.0	26.5	34.6	23.2	34.2	23.0
Treatment 2	45.9	34.3	23.4	15.0	50.8	34.2	35.2	23.1
ULCM with 50% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	3.8	14.8	13.4	**96.8**	12.0	13.6	**94.8**	**94.8**
Only treatment 1 is effective	9.0	**82.0**	0.0	0.0	**86.0**	**84.4**	3.2	3.2
Only treatment 2 is effective	10.0	0.0	**86.6**	3.2	0.0	0.2	1.6	1.6
Neither of treatment is effective	**77.2**	3.2	0.0	0.0	2.0	1.8	0.4	0.4
Patient allocation (n)	**Total N = 455**				**Total N = 428**			
Control	51.2	35.8	56.7	23.5	53.1	35.0	38.2	25.6
Treatment 1	48.4	24.7	56.7	27.0	37.3	25.0	37.0	25.0
Treatment 2	47.7	35.2	28.4	19.4	52.9	35.0	38.1	25.5
ULCM with 75% patients in Stage I								
Treatment selection (%)								
Both treatments are effective	1.8	12.6	11.6	**96.4**	11.4	12.6	**92.8**	**92.6**
Only treatment 1 is effective	10.0	**83.8**	0.0	0.0	**84.8**	**83.4**	3.4	3.4
Only treatment 2 is effective	10.4	0.0	**88.4**	3.6	0.2	0.4	3.0	3.2
Neither of treatment is effective	**77.8**	3.6	0.0	0.0	3.6	3.6	0.8	0.8
Patient allocation (n)	**Total N = 479**				**Total N = 463**			
Control	50.2	34.7	54.0	28.3	52.6	35.0	43.2	28.9
Treatment 1	49.3	29.1	54.5	30.1	42.7	28.8	43.1	28.9
Treatment 2	49.3	34.6	38.9	26.2	51.9	34.9	43.9	29.2

Table 6.4: The operating characteristics of the ULCM design including percentages of treatment selection, patients' allocation and mean sample sizes with different PFS distributions.

Designs	Subgroups				Subgroups			
	$g = 0$ $B_1 = 0$ $B_2 = 0$	$g = 1$ $B_1 = 1$ $B_2 = 0$	$g = 2$ $B_1 = 0$ $B_2 = 1$	$g = 3$ $B_1 = 1$ $B_2 = 1$	$g = 0$ $B_1 = 0$ $B_2 = 0$	$g = 1$ $B_1 = 1$ $B_2 = 0$	$g = 2$ $B_1 = 0$ $B_2 = 1$	$g = 3$ $B_1 = 1$ $B_2 = 1$
	Scenario 1				*Scenario 2*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
t=2	0.6	0.6	0.6	0.6	0.15	0.15	0.15	0.15
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
t=2	0.60	0.60	0.60	0.60	0.40	0.40	0.40	0.40
ULCM with PFS following exponential distribution								
Treatment selection (%)								
Both treatments are effective	**91.0**	**91.4**	**91.0**	**91.4**	1.8	2.0	2.0	1.8
Only treatment 1 is effective	3.4	3.6	3.4	3.6	8.8	7.4	9.0	7.6
Only treatment 2 is effective	3.6	3.6	3.8	3.4	9.2	8.4	7.2	7.4
Neither of treatment is effective	2.0	1.4	1.8	1.6	**80.2**	**82.2**	**81.8**	**83.2**
Patient allocation (n)	**Total N = 387**				**Total N = 487**			
Control	37.5	25.0	37.7	25.3	51.3	33.9	51.3	33.8
Treatment 1	39.9	26.3	39.1	25.9	46.5	32.0	47.5	31.3
Treatment 2	39.3	26.1	39.0	26.3	48.1	31.8	47.4	32.2
ULCM with PFS following gamma distribution								
Treatment selection (%)								
Both treatments are effective	**90.0**	**89.8**	**89.8**	**89.8**	2.4	3.0	2.8	3.0
Only treatment 1 is effective	3.8	4.0	4.0	4.0	7.4	8.6	7.0	8.6
Only treatment 2 is effective	3.4	3.4	3.4	3.4	6.8	6.0	5.6	5.4
Neither of treatment is effective	2.8	2.8	2.8	2.8	**83.4**	**82.4**	**84.6**	**83.0**
Patient allocation (n)	**Total N = 384**				**Total N = 483**			
Control	37.7	24.8	37.3	24.9	51.2	34.2	51.6	34.4
Treatment 1	38.7	25.5	39.0	25.9	47.0	30.8	47.0	31.0
Treatment 2	39.3	25.8	39.3	26.0	46.4	31.2	47.3	31.1
ULCM with PFS following piecewise-exponential distribution								
Treatment selection (%)								
Both treatments are effective	**96.6**	**97.0**	**97.2**	**97.2**	3.0	2.6	2.6	2.4
Only treatment 1 is effective	1.4	1.4	1.2	1.2	6.8	7.6	7.0	7.8
Only treatment 2 is effective	1.8	1.4	1.4	1.4	7.2	8.0	7.6	7.8
Neither of treatment is effective	0.2	0.2	0.2	0.2	**83.0**	**81.8**	**82.8**	**82.0**
Patient allocation (n)	**Total N = 361**				**Total N = 484**			
Control	36.2	24.0	35.9	24.0	50.7	33.7	50.6	34.2
Treatment 1	36.7	23.9	36.5	24.2	46.6	31.3	46.9	30.9
Treatment 2	35.9	23.7	35.6	23.9	47.8	31.7	47.4	32.2

Table 6.4 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 3				*Scenario 4*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.15	0.15	0.15	0.15
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.40	0.40	0.40	0.40
t=2	0.40	0.40	0.60	0.60	0.40	0.40	0.60	0.60
ULCM with PFS following exponential distribution								
Treatment selection (%)								
Both treatments are effective	6.0	19.2	12.0	**89.8**	2.8	2.2	10.0	9.4
Only treatment 1 is effective	7.4	**75.0**	0.4	4.4	7.8	7.2	0.0	0.0
Only treatment 2 is effective	13.0	0.2	**82.0**	4.2	12.0	13.4	**84.0**	**84.6**
Neither of treatment is effective	**73.6**	5.6	5.6	1.6	**77.4**	**77.2**	6.0	6.0
Patient allocation (n)	**Total N = 467**				**Total N = 486**			
Control	50.7	33.6	51.5	25.9	51.3	34.1	51.9	34.6
Treatment 1	47.9	27.4	52.2	27.1	47.5	31.5	50.7	33.7
Treatment 2	48.5	34.8	40.8	26.6	48.5	33.0	41.6	28.0
ULCM with PFS following gamma distribution								
Treatment selection (%)								
Both treatments are effective	2.0	10.6	8.0	**86.4**	1.6	2.0	7.6	7.0
Only treatment 1 is effective	5.8	**80.0**	0.0	4.2	5.6	5.0	0.0	0.0
Only treatment 2 is effective	8.0	0.2	**85.0**	6.6	7.4	8.4	**87.0**	**87.6**
Neither of treatment is effective	**84.2**	9.2	7.0	2.8	**85.4**	**84.6**	5.4	5.4
Patient allocation (n)	**Total N = 469**				**Total N = 483**			
Control	51.0	34.0	51.8	27.5	51.2	34.2	51.7	35.1
Treatment 1	47.7	28.2	50.0	27.8	45.5	30.5	48.9	32.5
Treatment 2	47.8	33.1	42.8	27.7	48.3	32.7	43.6	28.5
ULCM with PFS following piecewise-exponential distribution								
Treatment selection (%)								
Both treatments are effective	3.0	12.4	12.2	**93.8**	3.2	3.0	9.6	9.6
Only treatment 1 is effective	10.0	**83.2**	0.2	1.8	6.0	6.6	0.0	0.0
Only treatment 2 is effective	8.8	0.2	**85.0**	3.4	9.4	10.2	**88.0**	**88.0**
Neither of treatment is effective	**78.2**	4.2	2.6	1.0	**81.4**	**80.2**	2.4	2.4
Patient allocation (n)	**Total N = 467**				**Total N = 483**			
Control	50.2	35.4	53.8	25.3	51.2	33.7	53.4	35.7
Treatment 1	48.3	26.1	54.4	26.2	46.8	31.0	51.0	34.4
Treatment 2	48.7	36.2	37.4	24.7	48.4	32.7	39.0	25.9

Table 6.4 continues.

Designs	Subgroups				Subgroups			
	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1	g = 0 B_1 = 0 B_2 = 0	g = 1 B_1 = 1 B_2 = 0	g = 2 B_1 = 0 B_2 = 1	g = 3 B_1 = 1 B_2 = 1
	Scenario 5				*Scenario 6*			
p_{gt}								
t=0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
t=1	0.15	0.6	0.15	0.6	0.6	0.6	0.6	0.6
t=2	0.15	0.15	0.6	0.6	0.15	0.15	0.6	0.6
$\bar{F}_{gt}$								
t=0	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
t=1	0.40	0.60	0.40	0.60	0.60	0.60	0.60	0.60
t=2	0.40	0.40	0.68	0.68	0.40	0.40	0.60	0.60
ULCM with PFS following exponential distribution								
Treatment selection (%)								
Both treatments are effective	4.4	16.0	13.6	**91.6**	14.8	14.6	**90.0**	**90.4**
Only treatment 1 is effective	10.4	**75.6**	0.0	0.0	**80.6**	**81.2**	5.4	5.4
Only treatment 2 is effective	10.6	0.0	**86.4**	8.4	0.4	0.2	3.0	2.6
Neither of treatment is effective	**74.6**	8.4	0.0	0.0	4.2	4.0	1.6	1.6
Patient allocation (n)	**Total N = 459**				**Total N = 438**			
Control	50.7	34.3	54.3	25.2	52.8	35.2	37.6	25.4
Treatment 1	47.6	28.3	55.5	30.2	39.4	26.2	38.7	26.3
Treatment 2	48.0	34.3	30.8	20.1	53.3	36.0	40.1	26.9
ULCM with PFS following gamma distribution								
Treatment selection (%)								
Both treatments are effective	2.2	10.8	7.2	**92.6**	12.2	13.4	**89.4**	**89.2**
Only treatment 1 is effective	5.0	**82.2**	0.0	0.4	**83.6**	**81.8**	6.2	6.0
Only treatment 2 is effective	8.0	0.0	**92.4**	7.0	0.0	0.0	2.2	2.4
Neither of treatment is effective	**84.8**	7.0	0.4	0.0	4.2	4.8	2.2	2.4
Patient allocation (n)	**Total N = 457**				**Total N = 443**			
Control	52.3	34.6	53.8	25.5	53.2	35.2	39.2	26.7
Treatment 1	47.1	28.2	53.3	30.0	40.1	27.1	40.0	26.7
Treatment 2	47.1	33.3	31.6	20.6	51.8	34.8	41.0	27.1
ULCM with PFS following piecewise-exponential distribution								
Treatment selection (%)								
Both treatments are effective	3.8	14.8	13.4	**96.8**	12.0	13.6	**94.8**	**94.8**
Only treatment 1 is effective	9.0	**82.0**	0.0	0.0	**86.0**	**84.4**	3.2	3.2
Only treatment 2 is effective	10.0	0.0	**86.6**	3.2	0.0	0.2	1.6	1.6
Neither of treatment is effective	**77.2**	3.2	0.0	0.0	2.0	1.8	0.4	0.4
Patient allocation (n)	**Total N = 455**				**Total N = 428**			
Control	51.2	35.8	56.7	23.5	53.1	35.0	38.2	25.6
Treatment 1	48.4	24.7	56.7	27.0	37.3	25.0	37.0	25.0
Treatment 2	47.7	35.2	28.4	19.4	52.9	35.0	38.1	25.5

References

Ballarini, N.M. *et al.* (2021) "Optimizing subgroup selection in two-stage adaptive enrichment and umbrella designs," *Statistics in Medicine* [Preprint]. Available at: https://doi.org/10.1002/sim.8949.

Barker, A.D. *et al.* (2009) "I-SPY 2: An Adaptive Breast Cancer Trial Design in the Setting of Neoadjuvant Chemotherapy," *Clinical Pharmacology & Therapeutics*, 86(1), pp. 97–100. Available at: https://doi.org/10.1038/clpt.2009.68.

Berry, D.A. (2015) "The Brave New World of clinical cancer research: Adaptive biomarker-driven trials integrating clinical practice with clinical research," *Molecular Oncology*, 9(5), pp. 951–959. Available at: https://doi.org/10.1016/j.molonc.2015.02.011.

Berry, D.A., Herbst, R.S. and Rubin, E.J. (2012) "Reports from the 2010 Clinical and Translational Cancer Research Think Tank Meeting: Design Strategies for Personalized Therapy Trials," *Clinical Cancer Research*, 18(3), pp. 638–644. Available at: https://doi.org/10.1158/1078-0432.ccr-11-2018.

Brunk, H.D. *et al.* (1973) "Statistical Inference under Order Restrictions (The Theory and Application of Isotonic Regression)," *International Statistical Review*, 41(3), p. 395. Available at: https://doi.org/10.2307/1402630.

Cheung, Y.K. and Chappell, R. (2000) "Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities," *Biometrics*, 56(4), pp. 1177–1182. Available at: https://doi.org/10.1111/j.0006-341x.2000.01177.x.

Chu, Y. and Yuan, Y. (2018a) “A Bayesian basket trial design using a calibrated Bayesian hierarchical model.,” *Clinical Trials*, 15(2), pp. 149–158. Available at: https://doi.org/10.1177/1740774518755122.

Chu, Y. and Yuan, Y. (2018b) “BLAST: Bayesian Latent Subgroup Design for Basket Trials Accounting for Patient Heterogeneity,” *Applied Statistics*, 67(3), pp. 723–740. Available at: https://doi.org/10.1111/rssc.12255.

Dean, N.E. *et al.* (2020) “Creating a Framework for Conducting Randomized Clinical Trials during Disease Outbreaks,” *The New England Journal of Medicine*, 382(14), pp. 1366–1369. Available at: https://doi.org/10.1056/nejmsb1905390.

Desai, S. *et al.* (2007) “Phase I Study of Oxaliplatin, Full-Dose Gemcitabine, and Concurrent Radiation Therapy in Pancreatic Cancer,” *Journal of Clinical Oncology*, 25(29), pp. 4587–4592. Available at: https://doi.org/10.1200/jco.2007.12.0592.

Dutton, P. and Holmes, J. (2018) “Single arm two-stage studies: Improved designs for molecularly targeted agents,” *Pharmaceutical Statistics*, 17(6), pp. 761–769. Available at: https://doi.org/10.1002/pst.1896.

Freidlin, B., McShane, L.M. and Korn, E.L. (2010) “Randomized Clinical Trials With Biomarkers: Design Issues,” *Journal of the National Cancer Institute*, 102(3), pp. 152–160. Available at: https://doi.org/10.1093/jnci/djp477.

Ghisoni, E. *et al.* (2021) “Late-onset and long-lasting immune-related adverse events from immune checkpoint-inhibitors: An overlooked aspect in immunotherapy,” *European Journal of Cancer*, 149, pp. 153–164. Available at: https://doi.org/10.1016/j.ejca.2021.03.010.

Guo, W. *et al.* (2017) "A Bayesian interval dose-finding design addressingOckham's razor: mTPI-2," *Contemporary Clinical Trials*, 58, pp. 23–33. Available at: https://doi.org/10.1016/j.cct.2017.04.006.

Ivanova, A., Flournoy, N. and Chung, Y. (2007) "Cumulative cohort design for dose-finding," *Journal of Statistical Planning and Inference*, 137(7), pp. 2316–2327. Available at: https://doi.org/10.1016/j.jspi.2006.07.009.

Ji, Y. *et al.* (2010) "A modified toxicity probability interval method for dose-finding trials," *Clinical Trials*, 7(6), pp. 653–663. Available at: https://doi.org/10.1177/1740774510382799.

Ji, Y. and Wang, S.-J. (2013) "Modified Toxicity Probability Interval Design: A Safer and More Reliable Method Than the 3 + 3 Design for Practical Phase I Trials," *Journal of Clinical Oncology*, 31(14), pp. 1785–1791. Available at: https://doi.org/10.1200/jco.2012.45.7903.

Jones, C. and Holmgren, E. (2007) "An adaptive Simon Two-Stage Design for Phase 2 studies of targeted therapies," *Contemporary Clinical Trials*, 28(5), pp. 654–661. Available at: https://doi.org/10.1016/j.cct.2007.02.008.

Kang, D.G. *et al.* (2021) "Hierarchical Bayesian clustering design of multiple biomarker subgroups (HCOMBS)," *Statistics in Medicine* [Preprint]. Available at: https://doi.org/10.1002/sim.8946.

Kim, E.S. *et al.* (2011) "The BATTLE Trial: Personalizing Therapy for Lung Cancer," *Cancer Discovery*, 1(1), pp. 44–53. Available at: https://doi.org/10.1158/2159-8274.cd-10-0010.

Kroeger, N. *et al.* (2014) "Survival Outcome and Treatment Response of Patients with Late Relapse from Renal Cell Carcinoma in the Era of Targeted Therapy," *European Urology*, 65(6), pp. 1086–1092. Available at: https://doi.org/10.1016/j.eururo.2013.07.031.

Lee, Jeeyun *et al.* (2019) "Tumor Genomic Profiling Guides Patients with Metastatic Gastric Cancer to Targeted Treatment: The VIKTORY Umbrella Trial," *Cancer Discovery*, 9(10), pp. 1388–1405. Available at: https://doi.org/10.1158/2159-8290.cd-19-0442.

Lee, Y.T., Tan, Y. and Oon, C.E. (2018) "Molecular targeted therapy: Treating cancer with specificity," *European Journal of Pharmacology*, 834, pp. 188–196. Available at: https://doi.org/10.1016/j.ejphar.2018.07.034.

Lin, R. and Yuan, Y. (2020) "Time-to-event model-assisted designs for dose-finding trials with delayed toxicity," *Biostatistics*, 21(4), pp. 807–824. Available at: https://doi.org/10.1093/biostatistics/kxz007.

Liu, M., Wang, S.-J. and Ji, Y. (2020) "The i3+3 design for phase I clinical trials," *Journal of Biopharmaceutical Statistics*, 30(2), pp. 294–304. Available at: https://doi.org/10.1080/10543406.2019.1636811.

Liu, S., Yin, G. and Yuan, Y. (2014) "Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity," *The Annals of Applied Statistics*, 7(4). Available at: https://doi.org/10.1214/13-aoas661.

Liu, S. and Yuan, Y. (2015) “Bayesian optimal interval designs for phase I clinical trials,” *Applied Statistics*, 64(3), pp. 507–523. Available at: https://doi.org/10.1111/rssc.12089.

Mandrekar, S.J. and Sargent, D.J. (2009) “Clinical Trial Designs for Predictive Biomarker Validation: Theoretical Considerations and Practical Challenges,” *Journal of Clinical Oncology*, 27(24), pp. 4027–4034. Available at: https://doi.org/10.1200/jco.2009.22.3701.

Muler, J.H. *et al.* (2004) “Phase I Trial Using a Time-to-Event Continual Reassessment Strategy for Dose Escalation of Cisplatin Combined With Gemcitabine and Radiation Therapy in Pancreatic Cancer,” *Journal of Clinical Oncology*, 22(2), pp. 238–243. Available at: https://doi.org/10.1200/jco.2004.03.129.

Mullard, A. (2015) “NCI-MATCH trial pushes cancer umbrella trial paradigm,” *Nature Reviews Drug Discovery*, 14(8), pp. 513–515. Available at: https://doi.org/10.1038/nrd4694.

O’Quigley, J., Pepe, M.S. and Fisher, L.D. (1990) “Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer,” *Biometrics*, 46(1), p. 33. Available at: https://doi.org/10.2307/2531628.

Ouma, L.O. *et al.* (2021) “Treatment allocation strategies for umbrella trials in the presence of multiple biomarkers: A comparison of methods,” *Pharmaceutical Statistics*, 20(6), pp. 990–1001. Available at: https://doi.org/10.1002/pst.2119.

Ouma, L.O. *et al.* (2022) “Design and analysis of umbrella trials: Where do we stand?,” *Frontiers in Medicine*, 9. Available at: https://doi.org/10.3389/fmed.2022.1037439.

Paoletti, X., O’Quigley, J. and Maccario, J. (2004) “Design efficiency in dose finding studies,” *Computational Statistics & Data Analysis*, 45(2), pp. 197–214. Available at: https://doi.org/10.1016/s0167-9473(02)00323-7.

Parashar, D. *et al.* (2016) “An optimal stratified Simon two-stage design,” *Pharmaceutical Statistics*, 15(4), pp. 333–340. Available at: https://doi.org/10.1002/pst.1742.

Printz, C. (2010) “I-SPY 2 breast cancer clinical trial launches nationwide,” *Cancer* [Preprint]. Available at: https://doi.org/10.1002/cncr.25494.

Pusztai, L., Anderson, K.A. and Hess, K.R. (2007) “Pharmacogenomic Predictor Discovery in Phase II Clinical Trials for Breast Cancer,” *Clinical Cancer Research*, 13(20), pp. 6080–6086. Available at: https://doi.org/10.1158/1078-0432.ccr-07-0809.

Rao, A., Wiggins, C.L. and Lauer, R.T. (2018) “Survival outcomes for advanced kidney cancer patients in the era of targeted therapies,” *Annals of Translational Medicine*, 6(9), p. 165. Available at: https://doi.org/10.21037/atm.2018.04.44.

Ren, Y., Li, X. and Chen, C. (2021) “Statistical considerations of phase 3 umbrella trials allowing adding one treatment arm mid-trial,” *Contemporary Clinical Trials*, 109, p. 106538. Available at: https://doi.org/10.1016/j.cct.2021.106538.

Rogatko, A. *et al.* (2007) “Translation of Innovative Designs Into Phase I Trials,” *Journal of Clinical Oncology*, 25(31), pp. 4982–4986. Available at: https://doi.org/10.1200/jco.2007.12.1012.

Rudin, C.M. (2004) “Targeted cancer therapy,” *Nature*, 432(7015), pp. 294–297. Available at: https://doi.org/10.1038/nature03095.

Sargent, D.J. and Mandrekar, S.J. (2013) “Statistical issues in the validation of prognostic, predictive, and surrogate biomarkers,” *Clinical Trials* [Preprint]. Available at: https://doi.org/10.1177/1740774513497125.

Sawyers C. (2004) “Targeted cancer therapy,” *Nature*, 432(7015), pp. 294–297. Available at: https://doi.org/10.1038/nature03095.

Shan, M. *et al.* (2023) “Bayesian order constrained adaptive design for phase II clinical trials evaluating subgroup-specific treatment effect,” *Statistical Methods in Medical Research*, p. 096228022311587. Available at: https://doi.org/10.1177/09622802231158738.

Simon, R. and Roychowdhury, S. (2013) “Implementing personalized cancer genomics in clinical trials,” *Nature Reviews Drug Discovery*, 12(5), pp. 358–369. Available at: https://doi.org/10.1038/nrd3979.

Skolnik, J.M. *et al.* (2008) “Shortening the Timeline of Pediatric Phase I Trials: The Rolling Six Design,” *Journal of Clinical Oncology*, 26(2), pp. 190–195. Available at: https://doi.org/10.1200/jco.2007.12.7712.

Storer, B.E. (1989) “Design and Analysis of Phase I Clinical Trials,” *Biometrics*, 45(3), p. 925. Available at: https://doi.org/10.2307/2531693.

Storer, B.E. (2001) “An evaluation of phase I clinical trial designs in the continuous dose-response setting,” *Statistics in Medicine*, 20(16), pp. 2399–2408. Available at: https://doi.org/10.1002/sim.903.

Tourneau, C.L., Lee, J.S.H. and Siu, L.L. (2009) "Dose Escalation Methods in Phase I Cancer Clinical Trials," *Journal of the National Cancer Institute*, 101(10), pp. 708–720. Available at: https://doi.org/10.1093/jnci/djp079.

Wason, J. *et al.* (2015) "A Bayesian adaptive design for biomarker trials with linked treatments," *British Journal of Cancer*, 113(5), pp. 699–705. Available at: https://doi.org/10.1038/bjc.2015.278.

Woodcock, J. and LaVange, L.M. (2017) "Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both," *The New England Journal of Medicine*, 377(1), pp. 62–70. Available at: https://doi.org/10.1056/nejmra1510062.

Yan, F., Mandrekar, S.J. and Yuan, Y. (2017) "Keyboard: A Novel Bayesian Toxicity Probability Interval Design for Phase I Clinical Trials," *Clinical Cancer Research*, 23(15), pp. 3994–4003. Available at: https://doi.org/10.1158/1078-0432.ccr-17-0220.

Yang, S., Wang, S.-J. and Ji, Y. (2015) "An integrated dose-finding tool for phase I trials in oncology," *Contemporary Clinical Trials*, 45, pp. 426–434. Available at: https://doi.org/10.1016/j.cct.2015.09.019.

Yuan, Y. *et al.* (2018) "Time-to-Event Bayesian Optimal Interval Design to Accelerate Phase I Trials," *Clinical Cancer Research*, 24(20), pp. 4921–4930. Available at: https://doi.org/10.1158/1078-0432.ccr-18-0246.

Yuan, Y. and Yin, G. (2011) "Robust EM Continual Reassessment Method in Oncology Dose Finding," *Journal of the American Statistical Association*, 106(495), pp. 818–831. Available at: https://doi.org/10.1198/jasa.2011.ap09476.

Zang, Y. *et al.* (2019) "A Bayesian adaptive marker-stratified design for molecularly targeted agents with customized hierarchical modeling," *Statistics in Medicine*, 38(15), pp. 2883–2896. Available at: https://doi.org/10.1002/sim.8159.

Zhao, Y. *et al.* (2022) "A bayesian platform trial design to simultaneously evaluate multiple drugs in multiple indications with mixed endpoints," *Biometrics* [Preprint]. Available at: https://doi.org/10.1111/biom.13694.

Zhou, X. *et al.* (2008) "Bayesian adaptive design for targeted therapy development in lung cancer — a step toward personalized medicine," *Clinical Trials*, 5(3), pp. 181–193. Available at: https://doi.org/10.1177/1740774508091815.

Curriculum Vitae

Jiaying Guo

Education

- Ph.D. in Biostatistics, Indiana University, IUPUI, IN, Jul. 2023 (minor in Epidemiology)
- M.S. in Statistics, Georgia Institute of Technology, Atlanta, GA, Dec. 2009
- M.S. in Biology, Georgia Institute of Technology, Atlanta, GA, May 2009
- B.S. in Biology, Fudan University, Shanghai, P. R. China, Jul. 2005

Working Experiences

- Senior Principal Statistician, Eli Lilly and Company, Aug. 2014-present
- Senior Statistician, Syneos Health, Arp. 2010-Jul. 2014

Select Publications

- Fenske C, Boytsov N, **Guo J**, Dawson Z. Prescription Market Share and Treatment Patterns in Atopic Dermatitis: A Retrospective Observational Study Using US Insurance Claims. *Adv Ther.* 2022 May;39(5):2052-2064
- Edson-Heredia E, Anderson S, **Guo J**, Zhu B, Malatestinic WN, Wine-Lee L, Swanson E. Real-World Claims Analyses of Comorbidity Burden, Treatment Pattern, Healthcare Resource Utilization, and Costs in Pediatric Psoriasis. *Adv Ther.* 2021 Jul;38(7):3948-3961
- Foster SA, Zhu B, **Guo J**, Nikai E, Ojeh C, Malatestinic W, Goldblum O, Kornberg LJ, Wu JJ. Patient Characteristics, Health Care Resource Utilization, and Costs

Associated with Treatment-Regimen Failure with Biologics in the Treatment of Psoriasis. *J Manag Care Spec Pharm.* 2016 Apr;22(4):396-405

- Kimball AB, Edson-Heredia E, Zhu B, **Guo J**, Maeda-Chubachi T, Shen W, Bianchi MT. Understanding the Relationship Between Pruritus Severity and Work Productivity in Patients With Moderate-to-Severe Psoriasis: Sleep Problems Are a Mediating Factor. *J Drugs Dermatol.* 2016 Feb;15(2):183-8
- Foster S, Zhu B, **Guo J**, Ojeh C, Nikai E, Wu J. Treatment patterns associated with biologics in psoriasis patients. *J. Am. Acad. Dermatol.* 2015 May;72(5):AB256-AB256
- Edson-Heredia E, Zhu B, **Guo J**, Maeda-Chubachi T, Lebwohl M. Disease burden and quality of life in psoriasis patients with and without comorbid psoriatic arthritis: results from National Psoriasis Foundation panel surveys. *Cutis.* 2015 Mar;95(3):173-8
- Zhu B, Edson-Heredia E, **Guo J**, Maeda-Chubachi T, Shen W, Kimball AB. Itching is a significant problem and a mediator between disease severity and quality of life for patients with psoriasis: results from a randomized controlled trial. *Br J Dermatol.* 2014 Nov;171(5):1215-9
- Edson-Heredia E, Zhu B, **Guo J**, Lebwohl M. Comparison of patient reported outcomes by biologic treatment use in patients with moderate to severe plaque psoriasis: Results from a compiled dataset of National Psoriasis Foundation panel surveys. *J. Am. Acad. Dermatol.* 2013 Apr;68(4):AB195-AB195