8-2012

Bayesian Adaptive Designs for Early Phase Clinical Trials

Chunyan Cai

BAYESIAN ADAPTIVE DESIGNS FOR EARLY PHASE CLINICAL TRIALS

by

Chunyan Cai, B.S.

APPROVED:

_____

Supervisory Professor: Ying Yuan, Ph.D.

_____

Valen E Johnson, Ph.D.

_____

Yuan Ji, Ph.D.

_____

Yisheng Li, Ph.D.

_____

Edward Jackson, Ph.D.

APPROVED:

_____

Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

BAYESIAN ADAPTIVE DESIGNS FOR EARLY PHASE CLINICAL TRIALS

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
M. D. Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

by

Chunyan Cai, B.S.
Houston, Texas

August, 2012

To my dear parents

ACKNOWLEDGEMENTS

First of all, I would like to express my sincerest gratitude to my advisors, Dr. Ying Yuan and Dr. Valen E. Johnson for their guidance, advice and continuous support throughout my research work. I am thankful for their inspirations and encouragement that help to shape my technical, computational and writing skills for research. I greatly appreciate Dr. Ying's encouragement and advice in both research and life. I am also grateful to my supervisory committee members, Dr. Edward F. Jackson, Dr. Yuan Ji and Dr. Yisheng Li for their invaluable suggestions and comments on the revision of the thesis. I owe my thanks to my advisory committee members, Dr. Gary Rosner and Dr. Xin Shelley Wang, and thanks to my exam committee members, Dr. Jeffrey S. Morris, Dr. E. Neely Atkinson and Dr. Keith A. Baggerly for taking time to help with my presentations and qualify exams. Many thanks go in particular to Dr. Jing Ning and Dr. Yong Chen for their invaluable advice and encouragement in both research and life. I also would like to thank Dr. J. Jack Lee, Dr. Xuelin Huang, Dr. Li Zhang and other faculty members and staffs in the Division of Quantitative Sciences for their advice and great help through my PhD study. I also thank Dr. Victoria Knutson, Ms. Lourdes Perez and other GSBS staff members for their help and support in these years. My thanks also go to Ms. LeeAnn Chastain for her great help to improve the writing of this dissertation. I have also benefited from the IT team in the Division of Quantitative Sciences, who support the computing environment which is very helpful to produce the results and

this thesis. I also thank fellow students and colleagues in our department, Lin Yang, Lin Huo, Haiying Pang, Hua Li, Fei Jiang, Lixia Diao and others for their support, encouragement and friendship.

I could not have accomplished this thesis and this degree without the support from my family. I want to take this time to thank my parents, for their understanding and endless love in my pursuit of this degree and to all of my other endeavors in life. Thank my beloved one for his encouragement and support which carried me through difficult times. Finally, I want to thank all my beloved friends for their continuous support and faith in me.

ABSTRACT

BAYESIAN ADAPTIVE DESIGNS FOR EARLY PHASE CLINICAL TRIALS

Publication No. _____

Chunyan Cai, B.S.

Supervisory Professor: Ying Yuan, Ph.D.

My dissertation focuses mainly on Bayesian adaptive designs for phase I and phase II clinical trials. It includes three specific topics: (1) proposing a novel two-dimensional dose-finding algorithm for biological agents, (2) developing Bayesian adaptive screening designs to provide more efficient and ethical clinical trials, and (3) incorporating missing late-onset responses to make an early stopping decision.

Treating patients with novel biological agents is becoming a leading trend in oncology. Unlike cytotoxic agents, for which toxicity and efficacy monotonically increase with dose, biological agents may exhibit non-monotonic patterns in their dose-response relationships. Using a trial with two biological agents as an example, we propose a phase I/II trial design to identify the biologically optimal dose combination (BODC), which is defined as the dose combination of the two agents with the highest efficacy and tolerable toxicity. A change-point model is used to reflect the fact that the dose-toxicity surface of the combinational agents may plateau at higher dose levels, and a flexible logistic model is proposed to accommodate the possible non-monotonic pattern for the dose-efficacy relationship. During the trial, we continuously update the posterior estimates of toxicity and efficacy and assign patients to the most appro-

priate dose combination. We propose a novel dose-finding algorithm to encourage sufficient exploration of untried dose combinations in the two-dimensional space. Extensive simulation studies show that the proposed design has desirable operating characteristics in identifying the BODC under various patterns of dose-toxicity and dose-efficacy relationships.

Trials of combination therapies for the treatment of cancer are playing an increasingly important role in the battle against this disease. To more efficiently handle the large number of combination therapies that must be tested, we propose a novel Bayesian phase II adaptive screening design to simultaneously select among possible treatment combinations involving multiple agents. Our design is based on formulating the selection procedure as a Bayesian hypothesis testing problem in which the superiority of each treatment combination is equated to a single hypothesis. During the trial conduct, we use the current values of the posterior probabilities of all hypotheses to adaptively allocate patients to treatment combinations. Simulation studies show that the proposed design substantially outperforms the conventional multi-arm balanced factorial trial design. The proposed design yields a significantly higher probability for selecting the best treatment while at the same time allocating substantially more patients to efficacious treatments. The proposed design is most appropriate for the trials combining multiple agents and screening out the efficacious combination to be further investigated. The proposed Bayesian adaptive phase II screening design substantially outperformed the conventional complete factorial design. Our design allocates more patients to better treatments while at the same time providing higher power to identify the best treatment at the end of the trial.

Phase II trial studies usually are single-arm trials which are conducted to test the efficacy of experimental agents and decide whether agents are promising to be sent to phase III trials. Interim monitoring is employed to stop the trial early for futility to avoid assigning unacceptable number of patients to inferior treatments. We propose a Bayesian single-arm phase II design with continuous monitoring for estimating the response rate of the experimental drug. To address the issue of late-onset responses, we use a piece-wise exponential model to estimate the hazard function of time to response data and handle the missing responses using the multiple imputation approach. We evaluate the operating characteristics of the proposed method through extensive simulation studies. We show that the proposed method reduces the total length of the trial duration and yields desirable operating characteristics for different physician-specified lower bounds of response rate with different true response rates.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Dose-finding Design for Oncology Clinical Trials of Combinational Biological Agents

The paradigm of oncology drug development is expanding from traditional cytotoxic agents to biological agents [5, 17, 32]. Examples of biological agents include biospecimens targeting a specific tumor pathway, gene products aiming for DNA repair, and immunotherapies stimulating the immune system to attack a tumor. These novel agents differ from traditional cytotoxic agents in a variety of ways. For cytotoxic agents, toxicity and efficacy are typically assumed to monotonically increase with dose level. However, for biological agents, toxicity may increase at low dose levels and then approximately plateau at higher dose levels. For instance, when the toxicity of a biological agent is related to the inhibition of a biological pathway, the toxicity of the agent increases initially with dose since a high dose results in a high degree of inhibition; once the inhibition is saturated, the toxicity may be (or approximately) constant within a certain range of dose. In addition, the dose-efficacy curves for the biological agents may follow a non-monotonic pattern, and efficacy may even decrease at higher dose levels [17]. Therefore, traditional dose-finding designs with a focus on finding the maximum tolerated dose (MTD) [3, 35, 46, 59] are not suitable for trials of biological agents. Novel designs that consider both the toxicity

and efficacy of these agents are imperative.

Dose-finding designs that jointly model toxicity and efficacy are categorized as phase I/II designs. Numerous phase I/II designs have been proposed for traditional cytotoxic agents. Gooley et al. (1994) [16] proposed a phase I/II clinical trial in bone marrow transplantation to find a dose that balances the risks of two immunologic complications. Thall and Russell (1998) [49] proposed a phase I/II design to find a dose satisfying both safety and efficacy requirements based on a trinary outcome. O'Quigley, Hughes, and Fenton (2001) [33] presented a class of designs aiming to identify the dose yielding the highest treatment success rate for HIV studies. Thall and Cook (2004) [47] proposed a Bayesian phase I/II trial design based on trade-offs between toxicity and efficacy probabilities. Yin, Li, and Ji (2006) [58] developed a phase I/II Bayesian dose-finding design using toxicity and efficacy odds ratios. O'Quigley and Zohar (2006) [36] provided a comprehensive review of phase I/II designs. All of these phase I/II designs focus on single-agent trials and are not directly applicable to trials evaluating combinational agents.

For drug combination trials, a number of designs have been proposed to find the MTD of cytotoxic agents. Simon and Korn (1990) [42] described a mathematical model for the toxicity probability as a function of the weighted sum of the two doses to select cytotoxic drugs and dosages for a combination regimen. Afterwards, Korn and Simon (1993) [26] constructed a tolerable-dose diagram based on this simple mathematical model to guide the phase I trial design. Kramar, Lebecq and Candalh (1999) [27] proposed monotonically ordering of a selected subset of drug combinations which reduced the dose finding to a one-dimensional space. Thall et al. (2003) [48] developed a six-parameter logistic regression model of the toxicity probability to identify an entire "contour" of combinations. Conaway et al. (2004) [10] examined the simple

and partial orders for drug combinations based on the pool adjacent violators algorithm. Wang and Ivanova (2005) [53] proposed a two-stage Bayesian adaptive design to identify MTD combinations based on a logistic-type regression for toxicity probabilities. Yuan and Yin (2008) [60] proposed an adaptive two-dimensional dose-finding design that can accommodate any type of single-agent dose-finding method. They converted the two-dimensional dose-finding trial to a series of one-dimensional dose-finding subtrials and conducted the subtrials sequentially. Braun and Wang (2009) [6] proposed a hierarchical model for the dose-limiting toxicities (DLT) probability to identify MTD for novel combinations of cancer therapeutic agents which consider the subject heterogeneity for DLT. Recently, Wages, Conaway and O'Quigley (2011) [52] extended the continual reassessment method (CRM) to two-dimensional dose finding by converting a partially ordered two-dimensional dose space into a series of fully ordered dose sequences. All of these designs focus on phase I dose finding for cytotoxic agents and do not consider efficacy.

Published research on designs for phase I/II combination trials, in particular for biological agents, has been very limited. Yuan and Yin (2011) [61] developed a phase I/II design for drug combination trials, but that design focused on cytotoxic agents. Mandrekar, Cui and Sargent (2007) [31] proposed a novel phase I/II design for trials evaluating combinational biological agents based on a continuation ratio model for trinary outcomes (namely, "no response," "success" and "toxicity"). Our approach differs in several aspects: we model toxicity and efficacy as bivariate binary outcomes, use a change-point model to render the flexibility to consider that toxicity may plateau at high dose levels, and introduce a novel dose-finding algorithm to stochastically search the two-dimensional dose space, thereby encouraging the exploration of untried dose combinations.

Our research is motivated by a drug-combination trial at The University of Texas MD Anderson Cancer Center for patients diagnosed with relapsed lymphoma. The trial combined two novel biological agents, A and B (their names are masked to maintain confidentiality), that target two different components in the PI3K/AKT/mTOR signaling pathway. This pathway has been associated with several genetic aberrations related to the promotion of cancer [18]. Agent A is a PI3K kinase inhibitor and agent B is a downstream inhibitor of mTOR kinase within that pathway. Research has suggested that some types of lymphomas are promoted and maintained by the activation of the PI3K/AKT/mTOR pathway, making the pathway an important target for drug development [44]. Both agents A and B have individually demonstrated a partial inhibition of the pathway and some therapeutic activity. By combining these two agents, the investigators expect to obtain a more complete inhibition of the PI3k/AKT/mTOR pathway, and thereby to achieve better treatment responses. The trial investigates the combinations of 4 dose levels of agent A with 4 dose levels of agent B, which results in 16 dose combinations. The goal is to find the *biologically optimal dose combination* (BODC), defined as the dose combination with the highest efficacy and tolerable toxicity (e.g., with a toxicity probability < 0.4). The physicians expect the toxicity of the combinations to increase at low doses and become (approximately) flat at high doses, and they consider the possibility that the dose-efficacy curve of the combinations may be non-monotonic (i.e., the dose with the highest efficacy is not necessarily the highest dose).

We introduce a phase I/II design to identify the BODC for oncology trials of combinational biological agents. The proposed design explicitly accounts for the unique properties of biological agents. We propose a change-point model to reflect the property that the dose-toxicity surface of the combinational agents may plateau

at higher dose levels, and use a general logistic model with quadratic terms to accommodate the possible non-monotonic pattern of the dose-efficacy relationship. Our design is conducted in two stages: in stage I, we escalate doses along the diagonal of the dose combination matrix as a fast exploration of the dosing space; in stage II, based on the observed toxicity and efficacy data from stages I and II, we continuously update the posterior estimates of toxicity and efficacy and assign patients to the most appropriate dose combination. We propose a novel dose-finding algorithm to encourage sufficient exploration of the two-dimensional dose space, which facilitates the identification of the BODC. Extensive simulation studies show that the proposed design has desirable operating characteristics in identifying the BODC under various patterns of dose-toxicity and dose-efficacy relationships.

1.2 Screening Design for Combination Trials Combining Multiple Agents

The use of combination therapies [28, 38, 62] for cancer treatment can lead to treatment synergies that result in improved patient outcomes. The number of treatment combinations that must be tested is often quite large, however, which means that it is often not practical to conduct separate phase II trials on every possible combination of treatments. We describe a Bayesian adaptive trial design that facilitates the pooling of information obtained across treatment combinations by testing efficacy of all treatment combinations in a single trial. Important benefits of our trial designs include a reduction in the number of patients that must be recruited in order to evaluate each treatment combination, the assignment of a higher proportion of patients to efficacious treatments, faster patient accrual and more rapid completion of trials.

To motivate our design methodology, we consider a recent drug-combination

clinical trial conducted at MD Anderson Cancer Center to test the effectiveness of 16 combinations of 4 agents, A_1, A_2, A_3 and A_4, in reducing the symptom burden experienced by patients with late stage non-small cell lung cancer (NSCLC) who have received chemo-radiation therapies. The actual names of the drugs assessed in the trial are not specified here for reasons of confidentiality. The primary outcome variable for this trial was the area under the curve (AUC) for five symptoms (pain, fatigue, drowsiness, sleep disturbance and lack of appetite) measured daily using an interactive voice recording system (IVR) for 10 days following the onset of radiation therapy. Each symptom was measured using the MD Anderson Symptom Inventory (MDASI), which solicits ordinal ratings of symptoms on an 11-point scale ranging from 0 ("none at all") to 10 ("worst that can be imagined") [20]. In contrast to typical cancer-treating agents, which are generally cytotoxic, the combinations of four agents tested in this trial were known to have minimal risks of toxicity, and thus we focused herein on efficacy only. The goal of the trial is to identify the most efficacious combination to be further investigated in large trials.

A variety of screening designs have been proposed for use in trials of this general type. Among these, Thall, Simon and Ellenberg (1988) [50] proposed a two-stage phase II-III trial design to select the most promising treatment from k treatments in the first stage, and to compare the selected first stage treatment with the standard of care in the second stage. Schaid, Wieand and Therneau (1990) [41] presented a similar two-stage design that used survival data as an endpoint; that design allowed more than one treatment to be included in the second stage. Yao, Begg and Livingston (1996) [56] proposed a design to screen new treatments as a continuous process for identifying promising new therapeutic agents, and determined the optimal sample size to be used with their design. Yao and Venkatraman (1998) [57], and Wang and

Leung (1998) [54] extended that design to two-stage and fully sequential designs. Stout and Hardwick (2005) [45] developed a cost-based and constraint-based decision theoretic-approach to the design of screening trials. Rossell, Muller and Rosner (2007) [39] proposed a screening design based on Bayesian decision theoretics that uses optimal linear boundaries. Ding, Rosner and Muller (2008) [11] developed a more systematic decision-making optimal phase II screening trial design using a utility function that accounts for sampling costs and possible future payoff. However, none of these designs focus on screening combinations of multiple agents, a feature which is central to the designs we proposed in this work.

We model the main and synergistic effects of the treatment agents using a linear model, which facilitates borrowing information across the combinations. We cast the screening problem into a Bayesian hypothesis testing problem by constructing a series of hypotheses, each of which appoints one of the combinations as the most efficacious treatment. We utilize an encompassing prior with non-local prior constraints [25, 21] to accommodate the complex parameter constraints imposed by the hypotheses. During the trial conduct, based on the observed data, we continuously update the posterior probabilities of the hypotheses and use them to adaptively allocate patients to effective combinations and select the best treatment. Extensive simulation studies show that, compared to the standard (multi-arm) balanced factorial design, the proposed design yields a significantly higher probability of selecting the best treatment. It also allocates more patients to efficacious treatments.

1.3 Interim Monitoring for Late-onset Responses

Phase II trial studies usually are single-arm trials which are conducted to test the efficacy of experimental agents and decide whether agents are sufficiently

promising to be sent to phase III trials. To avoid assigning unacceptable number of patients to inferior treatments, interim monitoring is employed to stop the trial early for futility if there is sufficient evidence to determine the inefficiency of experimental agents. Many phase II trial designs with interim monitoring are proposed to evaluate the efficacy of the experimental agents[29]. Simon (1988) [43] presented a optimal two-stage design which minimizes the expected sample size. The early stopping criteria is applied to make an early stopping decision for futility at the end of first stage. Thall, Simon and Estey (1996) [51] proposed a new flexible statistical strategy to continuously monitor both safety and efficacy in single-arm cancer clinical trials. Wathen et al. (2008) [55] proposed a Bayesian single-arm phase II design to account for heterogeneity between patient prognostic subgroups. The subgroup-specific early-stopping rules are applied to allow terminate some subgroups and continue others. Johnson and Cook (2009) [19] derived a new class of Bayesian designs based on formal hypothesis tests using nonlocal alternative prior densities with continuous monitoring.

In general, interim monitoring based on previous responses assumes that the outcome could be observed shortly after the initiation of treatment such that the outcomes of the patients enrolled in the trial have been completely observed by the time of interim monitoring. However, this assumption may not always hold in practice, for example the case of late-onset responses [4, 8, 9] which may occur long after the assignment of treatment. Combining with the fast accrual rate, this would result in large number of missing responses at the time of interim monitoring. To address such late-onset responses, one possible approach is to suspend the accrual and wait until the previously enrolled patients are fully followed. Obviously this approach utilizes all the information and provides a good estimate of response rate.

However, it leads to an infeasibly long trial and needs to suspend the trial frequently which is not practical and inconvenient for trial administration. If we do not suspend the accrual and assign a newly arriving patient to treatment immediately, those patients under treatment might not have completed the assessment. At the time of interim monitoring, the early stopping decision will only be made based on the fully observed data thus far. This approach is also problematic which often overestimates the response rate and terminates the trial unappropriate.

We propose a Bayesian single-arm phase II design with continuous monitoring for estimating the response rate of the experimental drug. To address the issue of unobserved responses at the decision making time, we propose an approach which is built on missing data methodology. Specifically, we treat the unobserved responses as missing data and apply standard methods to estimate the response rate. We propose a piece-wise exponential model to estimate the hazard function of time to response data and handle the missing responses using the multiple imputation approach. For the proposed methods, we do not need to suspend patient accrual to wait for the full observation of the outcomes of patients under treatment. We evaluate the operating characteristics of the proposed method through extensive simulation studies. We show that the proposed method reduces the total length of the trial duration and yields a desirable operating characteristics for different physician-specified lower bounds of response rate with different true response rates.

CHAPTER 2

A Bayesian Phase I/II Design for Oncology Clinical Trials of Combinational Biological Agents

In this chapter, we introduce a phase I/II design to identify the BODC for oncology trials combining biological agents. Biological agents are playing an increasingly important role in oncology drug development. There are some unique features for the biological agents. The toxicity of biological agents is usually tolerable within the therapeutic dose range and may plateau at higher dose levels. In addition, the dose-efficacy curves for these agents often follow a non-monotonic pattern in which efficacy may decrease at higher dose levels. For cytotoxic agents, toxicity and efficacy are typically assumed to monotonically increase with dose level. Therefore, traditional dose-finding designs with a focus on finding the MTD are not suitable for trials of biological agents. Novel designs that consider both the toxicity and efficacy of these agents are in great demand.

We propose a dose-finding design that can explicitly account for the unique properties of biological agents. A change-point model is proposed to reflect the property that the dose-toxicity surface of the combinational agents may plateau at higher dose levels and a general logistic model with quadratic terms is applied to accommodate the possible non-monotonic pattern for the dose-efficacy relationship. Our design is conducted in two stages: in stage I, we escalate doses along the diagonal

of the dose combination matrix as a fast exploration of the dosing space; in stage II, based on the observed toxicity and efficacy data from stages I and II, we continuously update the posterior estimates of toxicity and efficacy and assign patients to the most appropriate dose combination. To encourage sufficient exploration of the two-dimensional dose space, we propose a novel dose-finding algorithm which facilitates the identification of the BODC. We conducted extensive simulation studies to evaluate the operating characteristics of our proposed design.

In following sections, we introduce the probability models and the phase I/II design for finding the BODC. We apply our design to the lymphoma clinical trial and examine the design's operating characteristics through extensive simulation studies.

2.1 Methods

2.1.1 Modeling Toxicity and Efficacy

Consider a trial combining J doses of biological agent A, denoted by $a_1 < a_2 < \cdots < a_J$, with K doses of biological agent B, denoted by $b_1 < b_2 < \cdots < b_K$. Without loss of generality, we assume $J \geq K$ and that the dose values of the a_j's and b_k's have been standardized to have mean 0 and standard deviation of 0.5. This standardization is used to anticipate the prior elicitation in Section 2.1.2. Let (a_j, b_k) denote the combination of dose a_j and dose b_k, and let p_{jk} and q_{jk} denote the toxicity and efficacy probabilities of (a_j, b_k), respectively, for $j = 1, 2, ..., J$, and $k = 1, 2, ..., K$. Here, toxicity and efficacy are two binary events that reflect the side effects (toxicity) and therapeutic effects (efficacy) of the biological agents. Therefore, p_{jk} and q_{jk} are simply the probabilities of the toxicity event and efficacy event, respectively, at dose combination (a_j, b_k). Specifically, the toxicity probability indicates the probability that a subject experiences dose-limiting toxicity and the efficacy probability represents the probability that there exists a direct or surrogate

marker of efficacy. The efficacy event can be tumor shrinkage or pathological response given by clinicians. The goal of the trial is to identify the BODC in the $J \times K$ dose combination matrix.

A change-point model for toxicity

Unlike cytotoxic agents, for which toxicity typically is assumed to monotonically increase with the dose level, the toxicity of biological agents may initially increase at low doses and then plateau at high doses. To accommodate this property of biological agents, we describe p_{jk}, the toxicity probability of (a_j, b_k), using a change-point model

$$\text{(2.1)}\quad \text{logit}(p_{jk}) = (\beta_0+\beta_1 a_j+\beta_2 b_k)I(\beta_0+\beta_1 a_j+\beta_2 b_k \leq \omega)+\omega I(\beta_0+\beta_1 a_j+\beta_2 b_k > \omega),$$

where $I(\cdot)$ is the indicator function and $(\beta_0, \beta_1, \beta_2,\ \omega)$ are unknown parameters. Under this model, the shape of the dose-toxicity surface initially is monotonic with the dose level but changes to flat once it passes the threshold defined by $\beta_0 + \beta_1 a_j + \beta_2 b_k = \omega$ (see Figure 2.1). We assume that $\beta_1 > 0$ and $\beta_2 > 0$ such that the toxicity probability initially increases with the doses of A and B before it plateaus, at which time the toxicity probability is given by $e^{\omega}/(1 + e^{\omega})$. We choose the change-point model for the dose-toxicity surface because of its intuitive interpretation and the ability to capture the threshold effect that may occur in some biological agents. Nevertheless, the choice of the toxicity model could be flexible as long as the model is able to accommodate the non-monotonic dose-toxicity relationship. For example, an alternative model $p_{jk} = \omega * \text{logit}^{-1}(\beta_0 + \beta_1 a_j + \beta_2 b_k)$ can also provide a good fit and yield good operating characteristics (results not shown).

In model (2.1), we did not include an interactive effect for the two agents (e.g., an interaction term $\beta_3 a_j b_k$) because the reliable estimation of such an interac-

tion term requires a large sample size (e.g., a few hundreds), which is typically not available in phase I trials. Note that for the purpose of dose finding, we do not seek to model the entire dose-toxicity surface but aim to obtain an adequate local fit to facilitate dose escalation and de-escalation. A model may provide a poor global fit to the entire dose-toxicity surface; however, as long as the model provides a good local fit around the current dose, it will lead to correct decisions of dose escalation and selection. O'Quigley and Paoletti (2003) [34] showed that simple parsimonious models often yield better operating characteristics than complex models for dose finding. In addition, In the context of drug combination trials, Wang and Ivanova (2005) [53] and Braun and Wang (2010) [6] found that a model without interaction performed as well as one with interaction for dose finding.

A second-order logit model for efficacy

For biological agents, the dose-efficacy curve often follows a non-monotonic pattern. For example, in immunotherapy trials, the dose-efficacy relationship could be bell-shaped. That is, the most effective dose may be a dose in the middle of the therapeutic dose ranges, and when a dose level is lower or higher than the most effective dose, efficacy decreases. To incorporate such a non-monotonic pattern for the dose-efficacy relationship, we assume that the efficacy probability of (a_j, b_k), that is, q_{jk}, follows a logistic model of the form

$$\text{logit}(q_{jk}) = \gamma_0 + \gamma_1 a_j + \gamma_2 b_k + \gamma_3 a_j^2 + \gamma_4 b_k^2, \tag{2.2}$$

where $(\gamma_0, \ldots, \gamma_4)$ are unknown parameters. The quadratic terms render the model adequate flexibility to capture the non-monotonic shape of the dose-efficacy surface. In this dose-efficacy model, we exclude the interaction effect $a_j b_k$ for the same reason described above.

2.1.2 Likelihood and Prior Specification

Suppose that at a certain stage of the trial, among n_{jk} patients treated at the paired dose (a_j, b_k), x_{jk} and y_{jk} patients have experienced dose-limiting toxicity and efficacy, respectively, where $j = 1, \cdots, J$ and $k = 1, \cdots, K$. Let $\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2\}$ and $\boldsymbol{\gamma} = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ denote the regression coefficients in models (2.1) and (2.2). The likelihood function of the observed data $\mathcal{D} = \{x_{jk}, y_{jk}\}$ can be expressed as

$$L(\mathcal{D}|\omega, \boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \prod_{j=1}^{J} \prod_{k=1}^{K} p_{jk}^{x_{jk}} (1 - p_{jk})^{n_{jk} - x_{jk}} \times q_{jk}^{y_{jk}} (1 - q_{jk})^{n_{jk} - y_{jk}}.$$

Let $f(\omega)$, $f(\boldsymbol{\beta})$, and $f(\boldsymbol{\gamma})$ denote the prior distributions for ω, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, respectively. Assuming prior independence among ω, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, we write the joint posterior distribution as

$$f(\omega, \boldsymbol{\beta}, \boldsymbol{\gamma}|\mathcal{D}) \propto L(\mathcal{D}|\omega, \boldsymbol{\beta}, \boldsymbol{\gamma}) f(\omega) f(\boldsymbol{\beta}) f(\boldsymbol{\gamma}),$$

from which the full conditional distributions can be obtained. The Gibbs sampler [12, 7] is used to obtain posterior draws of unknown parameters for statistical inferences.

For the prior specification of the efficacy model, we assign $\boldsymbol{\gamma}$ a weakly informative default prior $f(\boldsymbol{\gamma})$ proposed by [13] for logistic regression. To use this default prior, we first scale the actual values of the clinical doses to standardized values $\{a_j\}$ and $\{b_k\}$, which have mean 0 and standard deviation 0.5, and then assign an independent Cauchy distribution with center 0 and scale 2.5, Cauchy(0, 2.5), to the regression coefficients $\gamma_1, \cdots, \gamma_4$, and a Cauchy distribution with center 0 and scale 10, Cauchy(0,10), to the intercept γ_0. The advantages of using the weakly informative priors include (1) these priors are diffuse and provide reasonable coverage of the plausible values of the parameters, (for example, the prior Cauchy(0, 10) for the intercept expects the efficacy probability for an average case to be between 10^{-9} and $1 - 10^{-9}$); and (2) these priors are also appropriately regularized such that a

dramatic change in efficacy probability (e.g., from 0.01 to 0.5) is unlikely when dose changes by one level, which substantially improves the estimation stability while still being vague enough to ensure that the data are able to dominate the priors [13]. For the toxicity model, we use the default prior Cauchy(0, 10) for intercept β_0. We assign β_1 and β_2 a gamma prior distribution with shape 0.5 and rate 0.5 to ensure the monotonicity before the dose-toxicity surface reaches the change line in model (2.1). To specify a prior for ω, we assume that the toxicity probability at the plateau is between 0.2 and 0.8, which corresponds to a value of ω ranging from -1.39 to 1.39. Thus, we assign ω a normal prior $N(0, 4)$, which provides sufficient coverage for all plausible toxicity probabilities at the plateau, given by $e^{\omega}/(1 + e^{\omega})$.

2.2 Trial design

The proposed phase I/II design consists of two stages. Stage I is a run-in period, in which the goal is to explore the dose-combination space quickly and collect preliminary data so that the proposed probability models can be reliably estimated in stage II for systematic dose finding. We start stage I of the design by treating the first cohort of patients at the lowest dose combination (a_1, b_1), and then escalate the dose along the diagonal of the dose combination matrix until we encounter a dose combination that violates the safety requirement

$$Pr(p_{jk} < \phi|\mathcal{D}) > \delta, \tag{2.3}$$

where ϕ denotes the target toxicity upper limit and δ is a prespecified safety cutoff. If the dose matrix is not square (i.e., $J > K$), we first escalate the dose along the diagonal from (a_1, b_1) to (a_2, b_2) and so on until we reach (a_K, b_K); thereafter, we escalate the dose by holding the dose level of B at K and increasing the dose level of A from (a_K, b_K) to (a_{K+1}, b_K) and so on until we reach the highest dose

combination (a_J, b_K). In stage I, only a small fraction of patients are enrolled into the trial and the observed data are sparse. Therefore, in this stage, we evaluate the safety requirement based on a simple beta-binomial model rather than the proposed change-point toxicity model. Specifically, we assume that the number of toxicities x_{jk} follows a binomial distribution $Bi(n_{jk}, p_{jk})$, and that the toxicity probability p_{jk} follows a beta distribution $Beta(\zeta, \xi)$ with two shape parameters ζ and ξ. To ensure that the data dominate the posterior distribution, we set ζ=0.1 and ξ=0.2. Under the beta-binomial model, $Pr(p_{jk} < \phi|\mathcal{D}) = \mathcal{B}(\phi|\zeta + x_{jk}, \xi + n_{jk} - x_{jk})$, where $\mathcal{B}(\cdot)$ is the cumulative density function for a beta distribution. In stage I we also collect efficacy data; however, these data will not be used to determine the dose escalation. The rationale is that in this initial phase, as long as the doses are safe, we should explore the two-dimensional dose space as quickly as possible to learn the dose-toxicity and dose-efficacy surfaces.

Whenever a dose combination (a_j, b_k) violates the safety requirement, i.e., $Pr(p_{jk} < \phi|\mathcal{D}) \leq \delta$, or we reach the highest dose combination (a_J, b_K), stage I is then complete and the trial moves on to stage II. For this stage of the trial we invoke the toxicity and efficacy models described in Section 2 for systematic dose finding. Stage II dose finding is highlighted by two features. First, the proposed algorithm encourages the exploration of untried dose combinations to avoid the problem of trapping in suboptimal doses, which is of particular concern for combinations of biological agents. Because of complex drug-drug interactions and non-monotonic dose-response patterns, the assumed (simple) dose-response model is not expected to estimate the true dose-response surface well, especially at the beginning of the trial when only a few observations are available. Consequently, the resulting estimates of efficacy and toxicity may substantially deviate from the truth, and the "optimal"

dose identified based on these estimates may actually be a suboptimal dose. In other words, the dose with the highest estimate of efficacy is not necessarily the one actually having the highest efficacy. By intentionally visiting untried dose combinations, the proposed method increases the chance of finding better combinations and avoids trapping in suboptimal doses. Second, we introduce a concept of *g-degree neighbor* and *g-degree admissible neighbor* to facilitate the dose finding on the two-dimensional space, the details of which we describe next.

Assume that the current dose combination is (a_j, b_k) and define *g-degree neighbors* of (a_j, b_k), denoted by $\mathcal{N}_g$, as dose combinations $\{(a_{j'}, b_{k'})\}$ whose dose levels are different from (a_j, b_k) no more than g levels, i.e., $\mathcal{N}_g = \{(a_{j'}, b_{k'}) : |j' - j| \leq g \text{ and } |k' - k| \leq g\}$. Note that the dose set of $\mathcal{N}_g$ includes the current dose combination itself. We further define a g-degree admissible dose set $\mathcal{A}_g = \{(a_{j'}, b_{k'}) : (a_{j'}, b_{k'}) \in N_g, Pr(p_{j'k'} < \phi_T|\mathcal{D}) > \delta\}$, which is a subset of the g-degree neighbors $\mathcal{N}_g$ satisfying the pre-specified safety requirement $Pr(p_{j'k'} < \phi_T|\mathcal{D}) > \delta$. That is, $\mathcal{A}_g$ contains the safe g-degree neighbors of the dose combination (a_j, b_k).

Let N denote the prespecified maximum sample size, N_1 denote the number of patients in stage I, and $N_2 = N - N_1$ be the total number of patients available for stage II. Then the proposed dose-finding algorithm for stage II is described as follows:

1. Based on the accumulated trial data, we determine the dose set $\mathcal{A}_{g^*}$, where $g^* = \min\{g : \mathcal{A}_g \neq \emptyset, g \geq 1\}$. That is, $\mathcal{A}_{g^*}$ is the nonempty admissible set with the smallest degree g^*. If $\mathcal{A}_{g^*}$ does not exist, i.e., all investigational doses violate the safety requirement, we terminate the trial.

2. In $\mathcal{A}_{g^*}$, we identify the combination (a_{j^*}, b_{k^*}) that has the highest posterior mean of efficacy rate $\hat{q}_{j^*k^*}$.

3. If combination (a_{j^*}, b_{k^*}) has not been used to treat any patient thus far, or all doses in $\mathcal{A}_{g^*}$ have been used to treat patients, we assign the next cohort of patients to (a_{j^*}, b_{k^*}). However, if (a_{j^*}, b_{k^*}) has been used to treat patients and there are some untried doses in $\mathcal{A}_{g^*}$, before we decide to assign the next cohort of patients to (a_{j^*}, b_{k^*}), we compare $\hat{q}_{j^*k^*}$ against the following threshold:

$$\hat{q}_{j^*k^*} > \left(\frac{N_2 - n_2}{N_2}\right)^{\alpha}, \tag{2.4}$$

where n_2 is the total number of patients that have been treated in stage II and α is a known tuning parameter controlling how stringent the threshold is. If the condition (2.4) is not satisfied, (a_{j^*}, b_{k^*}) will be excluded from the admissible set $\mathcal{A}_{g^*}$ and we return to step 2.

4. We continue the above steps until exhaustion of the sample size, and select as the BODC the dose combination with the highest value of $\hat{q}_{jk}$ and satisfying the safety requirement $Pr(p_{jk} < \phi|\mathcal{D}) > \delta$.

Remark 1: The threshold (2.4) plays a key role in adaptively encouraging the exploration of untried doses and avoiding the problem of trapping in suboptimal doses during dose finding. At the beginning of stage II, when patients have not yet been treated in that stage, i.e., $n_2 = 0$, the value of $\{(N_2 - n_2)/N_2\}^{\alpha}$ equals 1. Consequently, condition (2.4) disallows treating patients at a dose that has been used previously and supports the exploration of untried doses. This is a sensible action because at the beginning of stage II the efficacy estimate $\hat{q}_{jk}$ is of large variability, and we should give high priority to using new doses rather than putting too much faith in the point estimate $\hat{q}_{jk}$. Toward the end of the trial (i.e., $n_2 \approx N_2$), we have more precise estimates of $\hat{q}_{jk}$ based on the accumulated data. As $\{(N_2 - n_2)/N_2\}^{\alpha}$

approaches 0, we essentially assign incoming patients to the dose combination with the highest value of $\hat{q}_{jk}$ because condition (2.4) is almost always satisfied. In condition (2.4), the tuning parameter α controls how fast $\{(N_2 - n_2)/N_2\}^{\alpha}$ decays from 1 to 0. The value of α can be calibrated to obtain desirable operating characteristics.

We summarize both stages of the proposed design as follows.

The proposed algorithm for finding BODC. The trial starts with the treatment of the first cohort of patients at the lowest dose (a_1, b_1). Suppose that patients are being treated at dose (a_j, b_k). A dose is safe if $\Pr(p_{jk} < \phi|\mathcal{D}) > \delta$; otherwise, the dose is deemed toxic.

Stage I Run-in Period

I1 If dose (a_j, b_k) is safe, escalate the dose and treat the next cohort at (a_{j+1}, b_{k+1}). If $j = k = K$, escalate the dose to (a_{j+1}, b_K). If (a_1, b_1) is deemed toxic, terminate the trial.

I2 Stage I is complete when either dose (a_j, b_k) is deemed toxic or the highest dose combination (a_J, b_K) is reached. Stage II then starts.

Stage II Systematic Dose Finding

II1 Based on the observed data, identify $\mathcal{A}_{g^*}$ as the nonempty set of safe neighbors of (a_j, b_k) with minimum degree g^*. If $\mathcal{A}_{g^*}$ does not exist (i.e., all experimental doses are deemed toxic), terminate the trial.

II2 Among the doses in $\mathcal{A}_{g^*}$, identify the dose (a_{j^*}, b_{k^*}) with the highest posterior mean of efficacy $\hat{q}_{j^*k^*}$.

II3 (a) If $n_{j^*k^*} = 0$ or $n_{rs} \neq 0$ for all $(a_r, b_s) \in \mathcal{A}_{g^*}$, treat the next cohort at dose (a_{j^*}, b_{k^*}).

(b) Otherwise,
$$\begin{cases} \text{If } \hat{q}_{j^*k^*} > \left(\frac{N_2 - n_2}{N_2}\right)^{\alpha} & \text{treat the next cohort at } (a_{j^*}, b_{k^*}), \\ \text{If } \hat{q}_{j^*k^*} \leq \left(\frac{N_2 - n_2}{N_2}\right)^{\alpha} & \text{remove dose } (a_{j^*}, b_{k^*}) \text{ from } \mathcal{A}_{g^*} \\ & \text{and go to step II2.} \end{cases}$$

II4 Repeat steps II2-4 until exhaustion of the sample size. Select as the BODC the dose combination with the highest $\hat{q}_{jk}$ among all safe doses.

2.3 Numerical Studies

2.3.1 Operating Characteristics

We conducted extensive simulations to evaluate the operating characteristics of the proposed phase I/II design. Step II3 in our design encourages exploration of untried dose combinations when sample size is small. This is an important feature of the proposed dose-finding algorithm. To evaluate the impact of this feature, we compared the proposed design to a "greedy" design that is otherwise identical except that it always assigns patients to the dose with the highest estimate of efficacy. Technically, this means that the greedy design replaces the condition (2.4) with $\hat{q}_{j^*k^*} > 0$ so that the dose with the highest efficacy among admissible dose set $\mathcal{A}_{g^*}$ is always selected.

We also compared our design with the phase I/II combination trial design proposed by Mandrekar, Cui, and Sargent(2007) [31]. For convenience, we refer to the latter design as the MCS design. The MCS design converts toxicity and efficacy into a mutually exclusive trinary outcome (namely, "no efficacy and no toxicity," "efficacy without toxicity" and "toxicity") and then uses a continuation ratio model to describe the relationship between this trinary outcome and the dose. To conduct a trial, the MCS design continuously updates the posterior estimates of the model parameters based on the observed data and assigns patients to the dose combination with the highest estimate of the probability of efficacy without toxicity (i.e., the MCS design adopts a greedy dose-finding algorithm).

We considered trials combining two biological agents, A and B, with a maximum sample size of 45 patients and a cohort size of 3. We investigated 8 different dose-toxicity and dose-efficacy scenarios (see Table 2.1). The first four scenarios consider the 4×4 combination trials with 4 dose levels for both agents A and B,

which were (0.075, 0.15, 0.225, 0.3) and (0.08, 0.16, 0.24, 0.32), respectively. We set the toxicity upper limit $\phi = 0.3$. The last four scenarios were taken from the work of Mandrekar, Cui, and Sargent (2007) [31], which involves the analysis of 5×3 combination trials with 5 doses of agent A, (0.60, 0.75, 0.90, 1.05, 1.35), and 3 doses of agent B, (0.60, 0.90, 1.20). The toxicity upper limit was $\phi = 0.33$.

In the proposed design, we set the safety cutoff $\delta = 0.4$ and the tuning parameter $\alpha = 2$, and used 2,000 posterior samples of unknown parameters ω, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ to make inference after 1,000 burn-in iterations based on the adaptive rejection Metropolis sampling algorithm [14]. Under each scenario, we carried out 2,000 simulated trials for each of the designs. We used C++ to implement the proposed design; the simulation code is available upon request.

The simulation results under scenarios 1-4 are summarized in Table 2.2, including the selection percentage for each dose combination as the BODC and the percentage of patients allocated to each dose combination (shown as subscripts). In scenario 1, the dose-toxicity surface initially increases with the dose levels of agents A and B and then plateaus in the right upper corner of the dose combination matrix with a toxicity probability of 0.25; the dose-efficacy relationship is non-monotonic, characterized by efficacy monotonically increasing with agent A but not with agent B. The true BODC is (a_4, b_2). Among the three designs, the proposed design performs the best with the highest selection probability (31.0%) and allocates the highest percentage of patients (15.9%) to the target dose combination. The greedy design is often trapped in the doses on the diagonal since it does not encourage exploration of untried dose combinations. As a result, it incorrectly selects the dose combination (a_4, b_4) as the BODC with the highest percentage. Moreover, the greedy design only allocates 10.0% patients to the true BODC, which is more than 1/3 lower than the

Table 2.1: Eight dose-toxicity and dose-efficacy scenarios for the simulation studies. The target BODCs are bolded.

Scenario	Agent A	Agent B: Toxicity probability 1	2	3	4	Efficacy probability 1	2	3	4
1	4	.25	**.25**	.25	.25	.42	**.60**	.38	.32
	3	.15	.25	.25	.25	.19	.44	.20	.18
	2	.10	.25	.25	.25	.12	.29	.15	.10
	1	.05	.10	.15	.25	.05	.22	.10	.08
2	4	.25	.25	.25	.25	.10	.29	.29	.42
	3	.15	.25	.25	**.25**	.25	.35	.43	**.60**
	2	.10	.25	.25	.25	.12	.24	.32	.39
	1	.05	.10	.15	.25	.05	.14	.28	.32
3	4	.25	.25	.25	.25	.05	.12	.18	.26
	3	.15	.25	.25	.25	.10	.15	.25	.30
	2	.10	.25	.25	.25	.14	.18	.30	.43
	1	.05	.10	.15	**.25**	.23	.28	.42	**.60**
4	4	**.17**	.25	.45	.55	**.60**	.35	.32	.28
	3	.12	.16	.25	.43	.42	.30	.28	.25
	2	.08	.10	.19	.22	.35	.28	.22	.20
	1	.05	.08	.12	.18	.25	.23	.19	.16
5	5	.07	.09	**.11**		.48	.53	**.64**	
	4	.05	.07	.09		.29	.36	.51	
	3	.04	.06	.08		.19	.28	.45	
	2	.03	.05	.07		.13	.22	.40	
	1	.02	.04	.06		.10	.19	.38	
6	5	.51	.52	.53		.19	.28	.45	
	4	.41	.42	.43		.19	.28	.45	
	3	.31	.32	.34		.19	.28	.45	
	2	.16	.18	.19		.19	.28	.45	
	1	.11	.13	**.15**		.19	.28	**.45**	
7	5	.41	.42	.43		.15	.24	.42	
	4	.21	.22	.24		.43	.49	.61	
	3	.06	.08	**.10**		.62	.66	**.74**	
	2	.04	.05	.07		.43	.49	.61	
	1	.03	.05	.07		.24	.32	.48	
8	5	.80	.81	.81		.15	.19	.19	
	4	.51	.52	.53		.34	.41	.48	
	3	.36	.37	.38		.43	.49	.61	
	2	.21	.22	.24		.53	.58	.68	
	1	.06	.08	**.10**		.62	.66	**.74**	

proposed design. The MCS design does not perform well, selecting the true BODC only 9.2% of the times. Scenarios 2 and 3 share the same dose-toxicity surface as scenario 1, i.e., toxicity initially increases and then plateaus, but possesses different shapes of the dose-efficacy surface. In scenario 2, combination (a_3, b_4) has the highest efficacy and is the true BODC. Our proposed design identifies (a_3, b_4) with the highest selection percentage 33.1% and assigns 18.5% patients to that dose combination. The greedy and MCS designs identify the true BODC 17.9% and 14.5% of the times and assign only 9.3% and 9.4% of the patients to the target, respectively. In scenario 3, a monotonic dose-efficacy relationship is assumed for agent B but not for agent A and the highest dose combination (a_1, b_4) is the true BODC. The proposed design again outperforms the other two designs. Scenario 4 is constructed to examine the case in which only toxicity monotonically increases with dose, but not efficacy. The proposed design yields a selection percentage of 46.3%, which is higher than those of the greedy design (39.1%) and the MCS design (26.5%).

The simulation results for trials with 5×3 combinations are shown in Table 2.3, indexed as scenarios 5-8. In scenario 5, toxicity is negligible for all dose combinations and efficacy monotonically increases with dose. The greedy design exhibits the best performance. This is mainly due to the coincidence that the greedy design would first escalate from dose combination (a_1, b_1) to (a_3, b_3) along the diagonal, then escalate up to the dose combination (a_5, b_3) during the run-in period in stage I. Therefore, after the initial dose escalation the greedy design would quickly identify (a_5, b_3) as the most desirable dose without exploring off-diagonal untried doses. Nevertheless, the proposed design exhibits a better performance than the MCS design. In scenario 6, toxicity monotonically increases with doses of both agents A and B, whereas efficacy only increases with agent B and is not affected by agent A. The

selection percentage of the proposed design is lower than that of the MSC design by 6.2%, but higher than that of the greedy design. In scenario 7, the selection percentage of the proposed design is higher than that of the MCS design (37.2% versus 30.2%), and in scenario 8, the selection percentage of the MCS design is 9.6% higher than the proposed design. In addition to the eight scenarios shown in Tables 2.2 and 2.3, we also considered additional scenarios (see Table 2.4) with different shapes of dose-toxicity and dose-efficacy relationships. The simulation results are summarized in Table 2.5 and demonstrate that the proposed design performs consistently well.

Figure 2.1: Surface of the toxicity probabilities for combinational agents using the proposed change-point model. Toxicity initially increases with dose level and plateaus after reaching the change line.

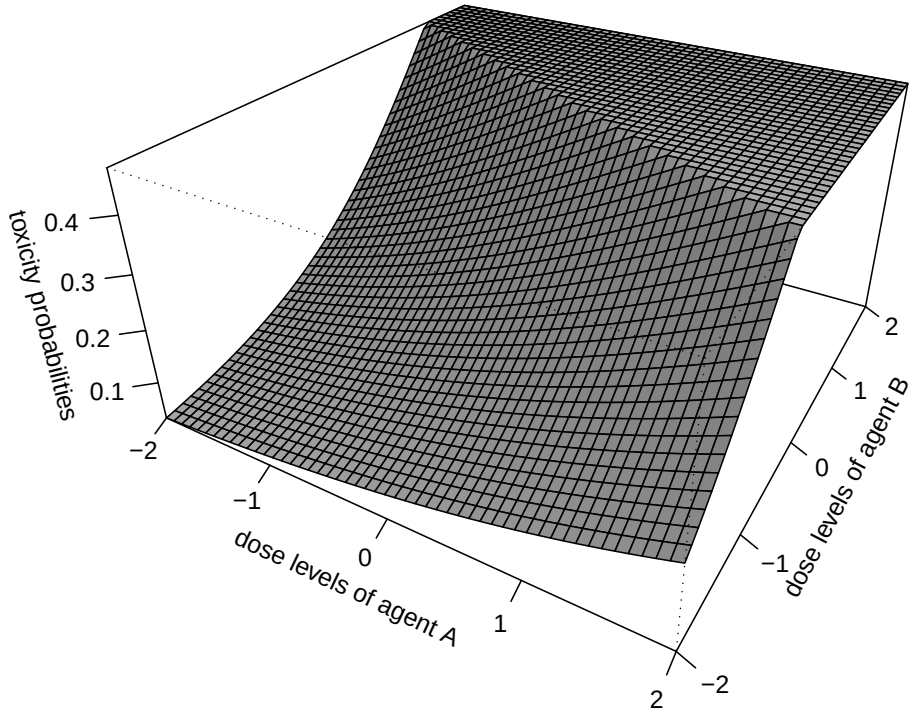

Table 2.2: The selection percentage and the percentage of patients treated at each dose combination (shown as the subscripts) for scenarios 1-4 under the proposed, greedy and MCS designs. The target BODCs are bolded.

		Agent B											
	Agent	Proposed Design				Greedy Design				MCS Design			
Scenario	A	1	2	3	4	1	2	3	4	1	2	3	4
1	4	$23.8_{14.1}$	$\mathbf{31.0_{15.9}}$	$10.8_{9.4}$	$8.9_{8.5}$	$18.2_{9.5}$	$\mathbf{21.5_{10.0}}$	$7.8_{5.3}$	$21.8_{26.5}$	$11.1_{5.0}$	$\mathbf{9.2_{5.0}}$	$11.1_{6.1}$	$29.9_{9.8}$
	3	$3.5_{3.9}$	$5.5_{6.0}$	$1.2_{6.9}$	$1.1_{4.6}$	$4.5_{3.0}$	$4.3_{3.0}$	$1.1_{9.5}$	$2.2_{3.2}$	$5.1_{4.7}$	$1.1_{1.8}$	$1.4_{3.2}$	$8.3_{7.0}$
	2	$0.9_{2.3}$	$2.7_{8.1}$	$0.8_{3.7}$	$0.5_{2.3}$	$1.2_{1.6}$	$4.2_{11.4}$	$0.9_{1.6}$	$0.6_{1.9}$	$0.6_{9.2}$	$0.2_{0.6}$	$0.8_{3.7}$	$8.4_{8.1}$
	1	$0.7_{7.6}$	$2.1_{2.8}$	$1.0_{2.1}$	$0.9_{1.8}$	$0.5_{8.4}$	$2.2_{1.9}$	$1.4_{2.1}$	$2.1_{1.2}$	$0.8_{9.3}$	$1.0_{9.4}$	$3.8_{9.0}$	$6.6_{8.0}$
2	4	$1.6_{2.1}$	$3.2_{3.2}$	$4.1_{6.4}$	$17.0_{13.7}$	$2.5_{1.6}$	$3.1_{2.3}$	$3.9_{3.7}$	$30.1_{32.0}$	$3.8_{3.3}$	$4.9_{3.4}$	$4.2_{3.6}$	$29.8_{9.8}$
	3	$2.5_{2.1}$	$2.8_{4.3}$	$7.1_{9.2}$	$\mathbf{33.1_{18.5}}$	$2.4_{2.3}$	$3.1_{2.3}$	$9.0_{13.9}$	$\mathbf{17.9_{9.3}}$	$2.6_{3.7}$	$1.1_{1.9}$	$1.4_{2.3}$	$\mathbf{14.5_{9.4}}$
	2	$0.7_{1.6}$	$1.5_{7.8}$	$3.4_{5.3}$	$9.6_{8.5}$	$0.8_{0.9}$	$1.1_{9.0}$	$3.0_{2.6}$	$8.2_{5.1}$	$0.4_{9.3}$	$0.6_{0.7}$	$1.2_{2.8}$	$14.9_{10.6}$
	1	$0.3_{7.3}$	$0.8_{1.6}$	$2.5_{2.7}$	$6.0_{5.7}$	$0.1_{7.7}$	$0.6_{0.9}$	$2.2_{2.3}$	$7.1_{3.9}$	$0.5_{9.4}$	$1.4_{9.6}$	$6.0_{9.8}$	$12.0_{10.3}$
3	4	$1.3_{2.1}$	$0.9_{2.5}$	$18_{4.6}$	$4.2_{6.9}$	$4.6_{1.2}$	$0.6_{1.9}$	$1.9_{3.0}$	$11.8_{20.7}$	$1.2_{1.6}$	$1.9_{1.7}$	$2.2_{2.3}$	$22.6_{8.5}$
	3	$0.8_{2.1}$	$0.7_{3.5}$	$1.2_{6.5}$	$4.3_{6.2}$	$0.7_{1.4}$	$0.7_{1.4}$	$1.2_{9.1}$	$3.5_{4.2}$	$1.0_{2.8}$	$0.4_{1.9}$	$1.4_{2.6}$	$11.1_{8.8}$
	2	$0.8_{2.8}$	$0.9_{7.9}$	$1.3_{3.9}$	$6.3_{6.0}$	$1.5_{1.6}$	$0.8_{9.8}$	$1.2_{1.6}$	$5.8_{4.2}$	$0.2_{9.3}$	$1.5_{1.3}$	$1.3_{3.3}$	$18.8_{11.9}$
	1	$3.8_{9.5}$	$3.8_{4.6}$	$9.4_{6.5}$	$\mathbf{54.0_{24.5}}$	$8.2_{17.0}$	$2.2_{2.3}$	$7.8_{4.7}$	$\mathbf{42.9_{16.0}}$	$0.5_{9.5}$	$1.7_{9.9}$	$9.9_{11.3}$	$\mathbf{23.8_{13.1}}$
4	4	$\mathbf{46.3_{18.9}}$	$6.8_{5.5}$	$3.4_{5.2}$	$1.3_{6.1}$	$\mathbf{39.1_{13.8}}$	$7.1_{5.2}$	$3.3_{3.6}$	$0.9_{9.8}$	$\mathbf{26.5_{10.0}}$	$25.6_{11.1}$	$11.3_{8.9}$	$5.2_{4.6}$
	3	$7.8_{5.5}$	$2.7_{5.0}$	$3.1_{8.6}$	$2.2_{4.5}$	$7.3_{3.9}$	$2.6_{2.9}$	$3.5_{13.2}$	$2.9_{3.9}$	$4.9_{6.8}$	$1.2_{2.6}$	$2.2_{4.7}$	$7.0_{5.6}$
	2	$5.3_{5.0}$	$1.9_{8.2}$	$1.5_{4.5}$	$3.1_{3.4}$	$3.9_{2.7}$	$3.0_{12.0}$	$1.8_{2.5}$	$3.9_{3.6}$	$0.4_{8.9}$	$0.0_{0.2}$	$0.6_{3.5}$	$9.2_{5.9}$
	1	$5.5_{10.2}$	$2.3_{3.6}$	$1.7_{2.7}$	$2.9_{3.0}$	$8.6_{16.1}$	$2.5_{2.0}$	$2.5_{1.8}$	$4.9_{2.9}$	$0.2_{8.5}$	$0.2_{8.6}$	$0.9_{5.6}$	$4.5_{4.6}$

Table 2.3: The selection percentage and the percentage of patients treated at each dose combination (shown as the subscripts) for scenarios 5-8 under the proposed, greedy and MCS designs. The target BODCs are bolded.

		Agent B								
	Agent	Proposed design			Greedy design			MCS design		
Scenario	A	1	2	3	1	2	3	1	2	3
5	5	$6.1_{7.8}$	$8.2_{10.7}$	$\mathbf{62.9_{34.0}}$	$2.8_{1.8}$	$3.1_{2.7}$	$\mathbf{70.9_{52.8}}$	$7.0_{2.6}$	$10.2_{4.9}$	$\mathbf{47.6_{15.5}}$
	4	$0.4_{1.3}$	$0.6_{4.9}$	$11.1_{11.3}$	$0.3_{0.2}$	$0.4_{0.2}$	$14.1_{16.5}$	$0.8_{1.1}$	$0.8_{2.3}$	$12.2_{9.9}$
	3	$0.1_{0.2}$	$0.4_{2.0}$	$4.7_{9.1}$	$0.2_{0.2}$	$0.1_{0.2}$	$5.1_{10.0}$	$0.0_{1.6}$	$0.4_{2.3}$	$9.6_{9.7}$
	2	$0.1_{0.2}$	$0.0_{6.7}$	$1.7_{2.4}$	$0.0_{0.0}$	$0.1_{6.9}$	$1.2_{1.1}$	$0.0_{9.2}$	$0.2_{2.6}$	$8.2_{10.1}$
	1	$0.1_{6.7}$	$0.3_{0.4}$	$3.4_{2.2}$	$0.1_{6.7}$	$0.2_{0.0}$	$1.2_{0.7}$	$0.0_{9.3}$	$0.4_{10.0}$	$2.4_{8.9}$
6	5	$1.7_{2.4}$	$1.0_{3.8}$	$3.6_{6.2}$	$3.2_{1.9}$	$1.3_{2.9}$	$3.1_{9.1}$	$0.6_{0.7}$	$0.2_{0.4}$	$0.0_{0.2}$
	4	$2.1_{3.6}$	$1.9_{5.5}$	$7.2_{9.8}$	$2.1_{3.1}$	$1.3_{3.1}$	$8.6_{14.4}$	$3.2_{3.2}$	$1.2_{1.3}$	$1.2_{1.1}$
	3	$2.3_{3.3}$	$1.8_{5.7}$	$16.2_{13.1}$	$2.5_{2.9}$	$2.3_{3.4}$	$17.2_{17.0}$	$5.4_{6.8}$	$1.2_{2.7}$	$6.0_{4.1}$
	2	$1.2_{3.6}$	$1.4_{8.1}$	$14.5_{8.8}$	$0.8_{2.4}$	$1.7_{9.6}$	$12.7_{7.7}$	$6.8_{14.4}$	$1.6_{4.2}$	$24.6_{11.2}$
	1	$1.7_{9.0}$	$3.5_{3.8}$	$\mathbf{30.0_{13.3}}$	$3.1_{10.6}$	$3.1_{2.2}$	$\mathbf{27.2_{9.8}}$	$3.4_{13.6}$	$7.2_{17.7}$	$\mathbf{36.2_{18.4}}$
7	5	$0.4_{0.4}$	$0.2_{2.0}$	$1.7_{7.6}$	$0.5_{0.2}$	$0.4_{0.78}$	$1.8_{9.6}$	$2.8_{3.2}$	$0.8_{1.8}$	$0.8_{1.7}$
	4	$1.2_{1.1}$	$2.7_{7.6}$	$23.8_{18.3}$	$0.8_{0.4}$	$1.1_{1.1}$	$26.6_{27.8}$	$8.8_{6.3}$	$4.4_{5.4}$	$7.0_{5.8}$
	3	$3.0_{1.8}$	$6.8_{7.8}$	$\mathbf{37.2_{23.0}}$	$2.0_{0.9}$	$3.2_{1.6}$	$\mathbf{41.4_{28.5}}$	$4.8_{5.3}$	$3.6_{4.7}$	$\mathbf{30.2_{11.4}}$
	2	$0.9_{1.1}$	$1.7_{7.8}$	$14.3_{10.3}$	$0.8_{0.7}$	$0.8_{7.8}$	$14.5_{9.8}$	$0.2_{9.6}$	$0.2_{3.7}$	$25.0_{12.3}$
	1	$0.4_{6.9}$	$0.5_{0.7}$	$4.6_{3.6}$	$0.5_{7.1}$	$0.4_{0.4}$	$4.9_{3.3}$	$0.2_{9.3}$	$1.2_{10.5}$	$8.0_{9.1}$
8	5	$0.2_{0.2}$	$0.1_{0.5}$	$0.1_{2.3}$	$0.2_{0.2}$	$0.0_{0.2}$	$0.0_{2.1}$	$0.0_{0.2}$	$0.0_{0.0}$	$0.0_{0.1}$
	4	$1.0_{1.2}$	$0.3_{2.8}$	$0.4_{4.9}$	$0.8_{0.9}$	$0.5_{1.6}$	$0.4_{6.0}$	$0.0_{1.6}$	$0.0_{0.3}$	$0.0_{0.1}$
	3	$1.9_{4.0}$	$1.9_{5.4}$	$3.0_{8.4}$	$2.1_{2.6}$	$1.2_{3.5}$	$3.4_{11.1}$	$3.4_{7.3}$	$0.6_{1.8}$	$0.0_{0.9}$
	2	$4.5_{7.2}$	$4.0_{10.3}$	$8.3_{6.5}$	$3.0_{3.9}$	$5.1_{14.2}$	$8.4_{6.5}$	$16.4_{20.0}$	$1.2_{4.0}$	$5.8_{4.5}$
	1	$20.5_{21.0}$	$13.0_{9.1}$	$\mathbf{34.2_{16.3}}$	$24.1_{26.0}$	$10.7_{5.8}$	$\mathbf{34.5_{15.3}}$	$6.6_{16.4}$	$19.8_{23.9}$	$\mathbf{43.8_{19.0}}$

Table 2.4: Additional dose-toxicity and dose-efficacy scenarios for the simulation studies. The target BODCs are bolded.

		Agent B							
Scenario	Agent	Toxicity probability				Efficacy probability			
	A	1	2	3	4	1	2	3	4
9	4	.14	.25	.25	.25	.19	.24	.39	.29
	3	.09	.15	**.18**	.25	.33	.45	**.60**	.42
	2	.04	.08	.13	.18	.14	.29	.37	.28
	1	.02	.04	.07	.12	.04	.17	.28	.17
10	4	.14	**.25**	.25	.25	.42	**.60**	.38	.32
	3	.09	.15	.18	.25	.19	.44	.20	.18
	2	.04	.08	.13	.18	.12	.29	.15	.10
	1	.02	.04	.07	.12	.05	.22	.10	.08
11	4	**.14**	.25	.25	.25	**.60**	.35	.32	.28
	3	.09	.15	.18	.25	.42	.30	.28	.25
	2	.04	.08	.13	.18	.35	.28	.22	.20
	1	.02	.04	.07	.12	.25	.23	.19	.16
12	4	.40	.40	.40	.40	.19	.24	.39	.29
	3	.15	.18	**.20**	.25	.33	.45	**.60**	.42
	2	.08	.12	.16	.20	.14	.29	.37	.28
	1	.01	.05	.12	.17	.04	.17	.28	.17
13	4	.40	.40	.40	.40	.10	.10	.18	.24
	3	.15	.18	.20	.25	.14	.14	.24	.43
	2	.08	.12	.16	**.20**	.23	.28	.42	**.60**
	1	.01	.05	.12	.17	.08	.10	.29	.42
14	4	.15	.18	.21	.25	.10	.10	.18	.24
	3	.10	.15	.19	.23	.14	.14	.24	.43
	2	.05	.12	.15	**.20**	.23	.28	.42	**.60**
	1	.01	.07	.12	.18	.08	.10	.29	.42
15	4	.15	**.18**	.21	.25	.42	**.60**	.38	.32
	3	.10	.15	.19	.23	.19	.44	.20	.18
	2	.05	.12	.15	.20	.12	.29	.15	.10
	1	.01	.07	.12	.18	.05	.22	.10	.08

Table 2.4 continued.

Scenario	Agent A	Agent B: Toxicity probability 1	2	3	4	Agent B: Efficacy probability 1	2	3	4
16	4	.15	.18	.21	.25	.19	.24	.39	.29
	3	.10	.15	**.19**	.23	.33	.45	**.60**	.42
	2	.05	.12	.15	.20	.14	.29	.37	.28
	1	.01	.07	.12	.18	.04	.17	.28	.17
17	4	.17	**.22**	.45	.50	.42	**.60**	.38	.32
	3	.12	.16	.25	.43	.19	.44	.20	.18
	2	.08	.10	.19	.22	.12	.29	.15	.10
	1	.05	.08	.12	.18	.05	.22	.10	.08
18	4	.17	.22	.45	.50	.10	.18	.24	.14
	3	.12	.16	.25	.43	.13	.24	.37	.26
	2	.08	.10	**.19**	.22	.24	.38	**.60**	.37
	1	.05	.08	.12	.18	.10	.23	.42	.22
19	4	.17	.22	.45	.50	.10	.10	.18	.24
	3	.12	.16	.25	.43	.14	.14	.24	.43
	2	.08	.10	.19	**.22**	.23	.28	.42	**.60**
	1	.05	.08	.12	.18	.08	.10	.29	.42
20	4	.06	.07	.08	.10	.24	.20	.15	.10
	3	**.05**	.06	.07	.08	**.60**	.43	.35	.25
	2	.04	.04	.05	.06	.39	.32	.24	.12
	1	.02	.03	.04	.05	.30	.20	.10	.05
21	4	.06	.07	.08	.10	.24	.18	.14	.10
	3	.05	.06	.07	.08	.37	.26	.20	.13
	2	**.04**	.04	.05	.06	**.60**	.37	.30	.24
	1	.02	.03	.04	.05	.42	.22	.15	.10
22	4	.25	.25	.25	.25	.24	.20	.15	.10
	3	**.15**	.25	.25	.25	**.60**	.43	.35	.25
	2	.10	.25	.25	.25	.39	.32	.24	.12
	1	.05	.10	.15	.25	.30	.20	.10	.05
23	4	.25	.25	.25	.25	.15	.10	.08	.05
	3	.15	.25	.25	.25	.24	.18	.14	.10
	2	.10	.25	.25	.25	.37	.26	.20	.13
	1	**.05**	.10	.15	.25	**.60**	.37	.30	.24

Table 2.5: The selection percentage and the percentage of patients treated at each dose combination (shown as the subscripts) for additional scenarios under the proposed, greedy and MCS designs. The target BODCs are bolded.

		Agent B											
	Agent	Proposed Design				Greedy Design				MCS Design			
Scenario	A	1	2	3	4	1	2	3	4	1	2	3	4
9	4	$1.5_{2.0}$	$4.5_{4.9}$	$12.8_{11.4}$	$8.7_{13.7}$	$0.9_{0.9}$	$2.9_{2.4}$	$12.6_{8.0}$	$13.1_{20.7}$	$1.2_{2.5}$	$5.3_{3.9}$	$9.4_{6.2}$	$53.9_{16.9}$
	3	$2.0_{1.6}$	$9.1_{8.1}$	$\mathbf{30.3_{17.0}}$	$17.6_{18.5}$	$0.8_{0.7}$	$5.5_{3.1}$	$\mathbf{39.1_{33.0}}$	$14.3_{9.1}$	$0.8_{2.7}$	$0.4_{1.2}$	$\mathbf{1.1_{2.5}}$	$14.9_{11.3}$
	2	$0.1_{0.4}$	$0.9_{7.4}$	$5.5_{6.9}$	$4.5_{8.5}$	$0.2_{0.2}$	$1.7_{8.5}$	$3.2_{2.2}$	$3.9_{2.9}$	$0.0_{8.9}$	$0.0_{0.1}$	$0.3_{1.5}$	$9.5_{10.1}$
	1	$0.1_{6.7}$	$0.2_{0.4}$	$0.9_{1.3}$	$1.0_{5.7}$	$0.1_{6.9}$	$0.1_{0.2}$	$0.4_{0.4}$	$1.1_{0.7}$	$0.0_{8.9}$	$0.2_{8.9}$	$0.2_{7.1}$	$2.3_{7.2}$
10	4	$25.7_{14.9}$	$\mathbf{37.2_{17.6}}$	$11.8_{10.7}$	$8.6_{25.7}$	$20.5_{9.8}$	$\mathbf{24.0_{10.2}}$	$6.6_{5.3}$	$28.2_{31.8}$	$3.0_{2.1}$	$\mathbf{7.6_{3.9}}$	$11.0_{7.3}$	$55.1_{17.1}$
	3	$1.4_{2.4}$	$4.5_{5.6}$	$2.1_{7.6}$	$1.4_{10.0}$	$1.5_{1.3}$	$3.8_{2.7}$	$1.4_{10.4}$	$2.0_{3.6}$	$0.5_{2.1}$	$0.4_{1.2}$	$1.0_{3.2}$	$11.2_{10.0}$
	2	$0.5_{1.3}$	$1.4_{7.3}$	$0.8_{3.6}$	$0.9_{4.4}$	$0.9_{0.9}$	$4.3_{10.2}$	$0.7_{1.3}$	$0.9_{1.8}$	$0.0_{9.0}$	$0.0_{0.0}$	$0.4_{2.3}$	$7.0_{9.3}$
	1	$0.6_{6.9}$	$1.3_{1.6}$	$0.6_{1.1}$	$1.0_{3.2}$	$0.4_{7.6}$	$2.1_{1.1}$	$0.5_{1.1}$	$1.9_{0.9}$	$0.0_{8.9}$	$0.0_{8.9}$	$0.2_{7.7}$	$2.4_{7.1}$
11	4	$\mathbf{43.1_{18.7}}$	$5.6_{5.3}$	$4.7_{6.9}$	$6.4_{8.7}$	$\mathbf{32.6_{10.9}}$	$5.3_{3.8}$	$4.9_{4.2}$	$12.5_{22.3}$	$\mathbf{11.1_{6.2}}$	$13.2_{8.1}$	$14.9_{10.5}$	$45.4_{15.3}$
	3	$7.4_{4.7}$	$1.9_{4.2}$	$3.1_{7.6}$	$4.3_{6.5}$	$6.0_{2.9}$	$2.1_{2.0}$	$3.3_{12.3}$	$4.3_{4.0}$	$0.8_{4.8}$	$0.5_{1.8}$	$1.0_{4.1}$	$7.1_{6.1}$
	2	$5.0_{4.2}$	$1.8_{7.6}$	$1.3_{4.2}$	$2.7_{3.8}$	$4.7_{2.5}$	$2.2_{11.4}$	$1.5_{1.6}$	$2.7_{2.7}$	$0.1_{9.0}$	$0.0_{0.1}$	$0.2_{2.8}$	$4.0_{4.9}$
	1	$5.2_{9.6}$	$2.4_{2.9}$	$1.6_{2.0}$	$3.2_{3.1}$	$8.1_{14.5}$	$1.6_{1.8}$	$2.3_{1.3}$	$5.8_{1.8}$	$0.1_{8.8}$	$0.0_{8.9}$	$0.1_{5.2}$	$1.2_{3.4}$
12	4	$1.7_{2.2}$	$3.5_{4.7}$	$7.2_{9.6}$	$3.6_{8.7}$	$1.8_{1.3}$	$3.8_{2.9}$	$6.5_{6.7}$	$4.0_{14.7}$	$5.1_{3.4}$	$6.2_{3.7}$	$8.9_{5.2}$	$24.7_{10.3}$
	3	$3.2_{2.5}$	$11.1_{8.9}$	$\mathbf{34.2_{18.1}}$	$16.3_{11.8}$	$2.2_{1.3}$	$8.1_{4.9}$	$\mathbf{43.5_{33.6}}$	$15.1_{9.6}$	$4.7_{3.7}$	$1.8_{2.1}$	$\mathbf{2.7_{3.2}}$	$22.2_{11.8}$
	2	$0.4_{1.1}$	$1.8_{7.8}$	$6.8_{7.6}$	$5.8_{5.8}$	$0.1_{0.4}$	$1.2_{8.5}$	$4.7_{3.3}$	$5.3_{3.6}$	$0.4_{8.6}$	$0.3_{0.4}$	$1.1_{2.5}$	$14.6_{11.5}$
	1	$0.2_{6.7}$	$0.6_{0.9}$	$1.9_{1.6}$	$1.3_{2.0}$	$0.0_{6.9}$	$0.4_{0.2}$	$1.1_{0.9}$	$1.6_{1.1}$	$0.0_{8.5}$	$0.1_{8.6}$	$2.1_{8.1}$	$5.1_{8.6}$
13	4	$1.2_{1.8}$	$0.9_{2.2}$	$0.8_{4.4}$	$2.6_{7.1}$	$2.1_{1.3}$	$1.0_{2.0}$	$1.7_{3.3}$	$4.8_{15.4}$	$3.0_{3.3}$	$4.0_{3.6}$	$3.4_{3.7}$	$20.8_{9.5}$
	3	$1.4_{2.0}$	$0.7_{3.8}$	$1.9_{7.8}$	$11.8_{10.0}$	$2.4_{2.0}$	$1.1_{2.5}$	$2.4_{11.8}$	$8.0_{7.1}$	$1.4_{4.7}$	$1.0_{3.6}$	$2.2_{3.7}$	$23.9_{12.0}$
	2	$2.4_{2.4}$	$2.1_{7.8}$	$5.3_{6.2}$	$\mathbf{37.9_{17.6}}$	$3.1_{2.5}$	$3.5_{10.9}$	$5.7_{3.8}$	$\mathbf{32.3_{12.7}}$	$0.2_{8.7}$	$0.4_{0.4}$	$1.0_{2.2}$	$\mathbf{21.8_{11.4}}$
	1	$0.7_{7.1}$	$0.3_{1.8}$	$4.3_{3.3}$	$25.5_{14.7}$	$1.1_{8.7}$	$0.9_{1.1}$	$4.2_{2.5}$	$25.5_{12.3}$	$0.0_{8.6}$	$0.5_{8.8}$	$3.5_{7.3}$	$12.6_{8.6}$

Table 2.5 continued.

		Agent B											
	Agent	Proposed Design				Greedy Design				MCS Design			
Scenario	A	1	2	3	4	1	2	3	4	1	2	3	4
14	4	$1.2_{1.8}$	$0.4_{2.2}$	$1.2_{4.9}$	$4.0_{8.2}$	$2.1_{1.8}$	$0.8_{6.0}$	$2.6_{3.3}$	$15.2_{22.9}$	$0.7_{2.7}$	$3.3_{3.7}$	$6.0_{5.9}$	$48.0_{15.6}$
	3	$1.4_{1.6}$	$0.8_{3.3}$	$1.7_{7.3}$	$12.6_{10.5}$	$1.2_{1.6}$	$0.7_{5.6}$	$2.6_{11.8}$	$7.0_{6.2}$	$0.5_{4.4}$	$0.4_{2.6}$	$1.2_{3.5}$	$18.4_{11.1}$
	2	$2.0_{2.0}$	$1.6_{7.6}$	$4.8_{5.6}$	$\mathbf{38.1_{18.0}}$	$3.0_{10.7}$	$3.9_{7.1}$	$4.6_{3.1}$	$\mathbf{27.7_{11.6}}$	$0.0_{8.9}$	$0.0_{0.1}$	$0.6_{1.7}$	$\mathbf{14.1_{9.4}}$
	1	$1.1_{7.1}$	$0.5_{1.6}$	$3.0_{3.3}$	$25.4_{14.9}$	$0.7_{1.1}$	$0.6_{0.7}$	$3.2_{2.0}$	$23.8_{11.6}$	$0.0_{8.9}$	$0.1_{9.0}$	$1.2_{6.0}$	$5.3_{6.4}$
15	4	$25.1_{14.8}$	$\mathbf{37.2_{17.2}}$	$11.6_{10.7}$	$9.0_{9.6}$	$18.9_{9.8}$	$\mathbf{22.9_{7.1}}$	$7.0_{5.8}$	$29.3_{32.2}$	$3.8_{2.5}$	$\mathbf{9.0_{4.5}}$	$11.6_{7.5}$	$52.6_{16.2}$
	3	$1.2_{2.5}$	$4.7_{5.8}$	$1.4_{7.4}$	$1.7_{5.4}$	$2.1_{2.7}$	$3.8_{5.8}$	$1.1_{10.4}$	$1.9_{3.3}$	$0.8_{2.6}$	$0.6_{1.6}$	$0.6_{3.2}$	$10.4_{9.3}$
	2	$0.3_{1.3}$	$2.1_{7.6}$	$0.4_{3.4}$	$0.8_{2.5}$	$0.9_{10.2}$	$3.9_{6.9}$	$0.8_{1.1}$	$1.3_{1.8}$	$0.1_{8.9}$	$0.0_{0.1}$	$0.5_{2.8}$	$6.9_{8.6}$
	1	$0.7_{6.9}$	$1.9_{1.8}$	$0.4_{1.3}$	$0.9_{1.8}$	$0.3_{1.3}$	$2.1_{0.7}$	$1.4_{1.3}$	$1.9_{0.9}$	$0.0_{8.8}$	$0.0_{8.9}$	$0.7_{7.7}$	$2.1_{6.8}$
16	4	$1.5_{2.0}$	$3.5_{5.1}$	$13.4_{11.6}$	$9.5_{11.1}$	$1.3_{2.2}$	$2.4_{6.2}$	$12.0_{7.8}$	$11.9_{20.4}$	$1.5_{2.4}$	$4.5_{3.9}$	$10.3_{6.6}$	$50.8_{15.7}$
	3	$2.2_{1.8}$	$8.2_{7.8}$	$\mathbf{29.4_{16.7}}$	$17.9_{13.1}$	$1.4_{3.3}$	$5.5_{4.5}$	$\mathbf{40.6_{33.3}}$	$13.9_{8.6}$	$0.8_{2.6}$	$0.4_{1.2}$	$\mathbf{1.5_{2.8}}$	$14.1_{10.6}$
	2	$0.3_{0.7}$	$1.1_{7.6}$	$6.3_{6.9}$	$3.6_{5.1}$	$0.3_{8.0}$	$1.4_{7.1}$	$3.4_{2.7}$	$4.1_{2.9}$	$0.0_{8.8}$	$0.0_{0.2}$	$0.7_{2.3}$	$9.8_{10.0}$
	1	$0.1_{6.7}$	$0.1_{0.7}$	$1.1_{1.3}$	$1.1_{2.0}$	$0.1_{0.2}$	$0.1_{0.4}$	$0.9_{0.7}$	$0.6_{0.9}$	$0.0_{8.8}$	$0.0_{8.9}$	$1.2_{7.6}$	$4.2_{7.6}$
17	4	$27.6_{15.2}$	$\mathbf{39.0_{17.4}}$	$6.8_{8.1}$	$2.2_{6.8}$	$30.5_{14.1}$	$\mathbf{32.8_{11.0}}$	$5.0_{6.1}$	$1.5_{13.2}$	$10.0_{4.2}$	$\mathbf{18.4_{6.5}}$	$12.2_{7.1}$	$9.4_{5.6}$
	3	$2.4_{2.9}$	$6.4_{6.6}$	$1.9_{8.4}$	$1.6_{5.0}$	$3.0_{4.5}$	$6.0_{7.0}$	$1.8_{12.3}$	$2.6_{4.1}$	$2.4_{3.4}$	$1.6_{2.0}$	$3.9_{4.4}$	$11.1_{9.0}$
	2	$0.4_{1.8}$	$2.1_{7.9}$	$0.9_{4.1}$	$1.4_{2.7}$	$0.8_{10.9}$	$4.6_{7.4}$	$1.3_{2.3}$	$1.6_{3.2}$	$0.4_{8.7}$	$0.2_{0.3}$	$2.2_{3.9}$	$16.4_{10.5}$
	1	$0.5_{7.2}$	$2.2_{2.0}$	$0.6_{1.8}$	$1.5_{2.0}$	$0.4_{1.6}$	$2.5_{1.6}$	$1.5_{1.6}$	$1.9_{1.6}$	$0.3_{8.5}$	$0.6_{8.6}$	$1.8_{8.3}$	$8.9_{9.1}$
18	4	$0.9_{1.6}$	$2.1_{2.9}$	$1.7_{5.0}$	$0.9_{6.1}$	$1.5_{3.1}$	$3.6_{9.2}$	$2.4_{3.4}$	$0.4_{7.6}$	$3.8_{3.2}$	$8.0_{4.5}$	$6.6_{5.1}$	$10.1_{5.8}$
	3	$0.9_{2.0}$	$2.7_{5.7}$	$9.8_{11.1}$	$2.7_{5.4}$	$1.5_{3.8}$	$3.5_{6.3}$	$13.6_{19.6}$	$3.1_{3.4}$	$0.8_{4.0}$	$1.4_{3.0}$	$4.0_{4.7}$	$17.8_{10.8}$
	2	$1.9_{2.9}$	$7.0_{10.0}$	$\mathbf{29.5_{14.9}}$	$12.7_{7.9}$	$2.7_{16.9}$	$11.1_{7.4}$	$\mathbf{22.0_{10.8}}$	$11.7_{6.3}$	$0.2_{8.5}$	$0.2_{0.4}$	$\mathbf{2.0_{2.9}}$	$25.1_{12.1}$
	1	$1.1_{7.5}$	$2.9_{3.4}$	$15.2_{8.1}$	$6.0_{5.4}$	$0.6_{1.8}$	$2.4_{1.3}$	$11.8_{5.2}$	$6.2_{4.9}$	$0.2_{8.4}$	$1.0_{8.7}$	$4.0_{7.9}$	$14.5_{10.0}$

Table 2.5 continued.

Scenario	Agent A	Agent B: Proposed Design 1	Proposed Design 2	Proposed Design 3	Proposed Design 4	Greedy Design 1	Greedy Design 2	Greedy Design 3	Greedy Design 4	MCS Design 1	MCS Design 2	MCS Design 3	MCS Design 4
19	4	$1.2_{1.8}$	$1.1_{2.5}$	$1.3_{4.3}$	$1.9_{6.5}$	$2.2_{2.0}$	$1.8_{2.7}$	$1.2_{3.2}$	$0.9_{10.6}$	$4.2_{4.2}$	$7.9_{5.1}$	$6.8_{5.6}$	$8.7_{5.5}$
	3	$1.5_{2.0}$	$0.5_{4.1}$	$2.2_{8.1}$	$6.5_{8.1}$	$2.0_{1.8}$	$1.2_{2.3}$	$2.5_{12.7}$	$5.1_{5.7}$	$1.2_{4.8}$	$0.8_{3.0}$	$2.7_{4.3}$	$18.9_{10.6}$
	2	$2.2_{2.5}$	$2.5_{7.9}$	$6.2_{6.8}$	$\mathbf{38.7_{16.9}}$	$3.6_{2.3}$	$4.0_{11.3}$	$6.6_{5.2}$	$\mathbf{33.0_{14.0}}$	$0.2_{8.5}$	$0.0_{0.3}$	$1.5_{2.6}$	$\mathbf{26.2_{11.5}}$
	1	$0.9_{7.4}$	$0.8_{2.0}$	$4.0_{3.8}$	$26.5_{15.1}$	$0.4_{8.6}$	$0.9_{1.4}$	$5.7_{2.9}$	$27.2_{13.3}$	$0.4_{8.6}$	$0.4_{8.8}$	$4.3_{7.2}$	$15.3_{9.2}$
20	4	$8.1_{5.3}$	$1.9_{2.4}$	$1.2_{4.9}$	$1.9_{7.1}$	$9.3_{4.4}$	$2.8_{2.2}$	$1.7_{2.2}$	$1.1_{10.0}$	$0.6_{4.7}$	$1.0_{5.6}$	$2.0_{8.8}$	$95.1_{27.7}$
	3	$\mathbf{36.2_{13.7}}$	$7.2_{7.1}$	$4.5_{8.2}$	$3.5_{5.5}$	$\mathbf{24.4_{7.6}}$	$8.9_{4.9}$	$6.7_{15.8}$	$3.9_{3.6}$	$\mathbf{0.0_{5.2}}$	$0.0_{1.0}$	$0.1_{3.2}$	$1.0_{3.6}$
	2	$14.9_{9.3}$	$5.1_{9.1}$	$1.9_{5.3}$	$1.0_{2.9}$	$8.1_{4.2}$	$8.3_{16.9}$	$3.5_{2.9}$	$2.1_{2.4}$	$0.0_{9.5}$	$0.0_{0.0}$	$0.0_{2.6}$	$0.0_{2.4}$
	1	$8.5_{11.3}$	$2.1_{4.2}$	$0.5_{2.0}$	$1.2_{1.8}$	$12.6_{18.0}$	$3.3_{2.4}$	$1.4_{1.3}$	$1.7_{1.1}$	$0.0_{9.6}$	$0.0_{9.6}$	$0.0_{4.4}$	$0.0_{1.9}$
21	4	$4.9_{3.6}$	$1.1_{1.8}$	$0.9_{4.2}$	$1.3_{6.7}$	$5.1_{2.7}$	$0.9_{1.3}$	$0.9_{1.6}$	$0.7_{8.9}$	$0.9_{5.1}$	$1.1_{6.0}$	$3.4_{9.2}$	$93.5_{27.0}$
	3	$11.0_{6.5}$	$1.7_{4.0}$	$0.9_{7.1}$	$0.5_{4.2}$	$11.1_{3.8}$	$1.6_{1.3}$	$1.3_{12.5}$	$0.4_{1.6}$	$0.0_{7.2}$	$0.0_{2.7}$	$0.0_{4.3}$	$0.5_{3.5}$
	2	$\mathbf{37.5_{17.4}}$	$3.5_{8.5}$	$1.5_{4.2}$	$1.8_{2.2}$	$\mathbf{17.1_{6.2}}$	$5.7_{15.4}$	$2.2_{1.6}$	$3.1_{2.4}$	$\mathbf{0.0_{9.5}}$	$0.0_{0.0}$	$0.0_{1.8}$	$0.2_{1.1}$
	1	$27.1_{20.5}$	$2.8_{4.9}$	$1.4_{2.2}$	$1.9_{2.0}$	$43.6_{36.3}$	$2.4_{1.8}$	$1.2_{1.1}$	$2.6_{1.6}$	$0.0_{9.5}$	$0.0_{9.5}$	$0.2_{2.7}$	$0.2_{0.9}$
22	4	$7.7_{5.8}$	$2.1_{2.6}$	$1.6_{3.7}$	$0.9_{5.6}$	$8.2_{5.1}$	$1.9_{2.1}$	$1.4_{2.1}$	$0.9_{7.4}$	$16.6_{9.7}$	$15.2_{9.4}$	$14.5_{9.4}$	$25.6_{9.0}$
	3	$\mathbf{35.4_{15.3}}$	$7.2_{7.0}$	$3.9_{7.7}$	$2.5_{4.4}$	$\mathbf{29.9_{11.1}}$	$8.2_{5.1}$	$5.2_{13.1}$	$3.5_{3.2}$	$\mathbf{10.0_{8.7}}$	$0.8_{2.4}$	$1.7_{4.2}$	$4.2_{3.7}$
	2	$16.2_{10.4}$	$3.5_{9.3}$	$1.8_{5.1}$	$1.0_{2.6}$	$7.5_{5.1}$	$7.0_{16.1}$	$2.5_{3.0}$	$1.2_{1.8}$	$2.0_{9.5}$	$0.4_{0.6}$	$1.0_{3.7}$	$4.0_{3.3}$
	1	$7.6_{12.1}$	$2.4_{4.4}$	$0.5_{2.6}$	$1.1_{1.6}$	$12.4_{19.8}$	$2.8_{2.5}$	$0.9_{1.6}$	$1.8_{0.9}$	$0.8_{9.1}$	$0.2_{8.9}$	$1.1_{5.6}$	$1.2_{2.9}$
23	4	$1.5_{1.8}$	$0.1_{0.9}$	$0.2_{2.8}$	$0.4_{5.0}$	$1.1_{0.9}$	$0.1_{0.5}$	$0.1_{0.7}$	$0.2_{5.6}$	$8.3_{6.0}$	$8.0_{5.6}$	$8.8_{6.1}$	$23.0_{7.4}$
	3	$0.9_{2.8}$	$0.4_{3.4}$	$0.1_{6.0}$	$0.1_{2.8}$	$1.1_{1.2}$	$0.2_{0.5}$	$0.2_{9.8}$	$0.1_{0.7}$	$4.5_{6.8}$	$1.6_{3.4}$	$1.6_{5.3}$	$6.7_{5.4}$
	2	$5.5_{8.0}$	$0.8_{7.6}$	$0.2_{3.4}$	$0.1_{1.1}$	$1.7_{1.4}$	$1.3_{12.8}$	$0.3_{0.5}$	$0.4_{0.7}$	$1.2_{8.8}$	$0.4_{0.7}$	$1.4_{4.2}$	$8.9_{6.4}$
	1	$\mathbf{73.4_{39.0}}$	$6.3_{8.5}$	$2.1_{3.7}$	$3.6_{3.2}$	$\mathbf{82.3_{60.5}}$	$2.4_{1.9}$	$1.6_{1.4}$	$1.7_{1.2}$	$\mathbf{1.8_{8.7}}$	$1.6_{8.8}$	$6.1_{8.6}$	$15.3_{7.9}$

CHAPTER 3

Bayesian Adaptive Phase II Screening Design for Combination Trials

In this chapter, we propose a Bayesian adaptive screening design for combination trials. There is an increasing trend to use the combination therapies for cancer treatment. Combination therapies can lead to treatment synergies that result in improved patient outcomes. Therefore, the number of treatment combinations that must be tested is often quite large. Conducting separate phase II trials on every possible combination of treatments is often not practical. Novel designs that can test the efficacy all combinations in a single trial are imperative.

Toward this goal, we describe a Bayesian adaptive trial design that facilitates the pooling of information obtained across treatment combinations. We model the main and synergistic effects of the treatment agents using a linear model, which facilitates borrowing information across the combinations. We cast the screening problem into a Bayesian hypothesis testing problem. We construct a series of hypotheses, each of which appoints one of the combinations as the most efficacious treatment. We utilize an encompassing prior with non-local constraints to accommodate the complex parameter constraints imposed by the hypotheses. During the trial conduct, based on the observed data, we continuously update the posterior probabilities of the hypotheses and use them to adaptively allocate patients to effec-

tive combinations and select the best treatment. We conduct extensive simulation studies to evaluate the performance of the proposed design. The comparison to the standard (multi-arm) balanced factorial design show that our proposed design selects the best treatment with a significantly higher probability and allocates more patients to efficacious treatments.

In following sections, we describe our model, prior specification and trial design. We examine the operating characteristics of our design using simulation studies.

3.1 Methods

3.1.1 Probability Model

We consider trials to evaluate the treatment effects of all possible combinations of k treatment agents, $A_1, A_2, \cdots$, and A_k. We assume that each drug combination is assigned to one treatment arm, although it is straightforward to extend our design to trials where some combinations are excluded. Given k agents, there are $\binom{k}{r}$ different r-agent combinations, $r = 0, 1, \cdots, k$, resulting in a total of $p = \sum_{r=0}^{k} \binom{k}{r} = 2^k$ combinations, including placebo group, to be evaluated. The goal of the trial is to identify the most efficacious treatment combination.

The outcome variable in the trial that motivates our research represents the mean change in the patient-reported symptom score. We assume that the outcome for the ith patient, y_i, is continuous and follows a linear model of the form

(3.1)

$$y_i = \beta_0 + \beta_1 I_i(A_1) + \beta_2 I_i(A_2) + \cdots + \beta_{1,2} I_i(A_1, A_2) + \cdots + \beta_{1,2,\cdots,k} I_i(A_1, A_2, \cdots, A_k) + \epsilon_i$$

where β_0 is the intercept of the linear model and $I_i(\cdot)$ is an indicator of whether patient i receives the given agents. For example, if patient i receives a combination

of A_1 and A_2, then $I_i(A_1) = I_i(A_2) = I_i(A_1, A_2) = 1$; whereas all the other indicator functions are then 0. Model (3.1) is flexible and accounts for the main and interaction effects of combining agents. Specifically, β_k represents the main treatment effect of A_k, $\beta_{k,k'}$ represents the two-way interaction or synergistic effect between A_k and $A_{k'}$ when $k \neq k'$, and so on. We assume that the residual ϵ_i follows a normal distribution with mean 0 and variance σ^2. Binary and time-to-event outcomes can be modeled using a similar linear structure within a generalized linear model framework.

To cast the problem into a hypothesis testing framework, we define the null hypothesis H_0 to assert that no treatment is better than the placebo, and a series of alternative hypotheses $H_1, \ldots, H_{p-1}$, where H_j asserts that the jth treatment combination is superior to all others. In our trial, for example, treatment j is superior to treatment k if it leads to a greater reduction in symptom burden. Let $\theta_0(\boldsymbol{\beta})$ denote the effect of the placebo and let $\theta_j(\boldsymbol{\beta})$, $j = 1, \ldots, p-1$, denote the net treatment effect of the jth combination (or treatment arm). Under the linear model (3.1), the treatment effect, $\theta_j(\boldsymbol{\beta})$, is a linear combination of the regression parameters, β's. For example, the treatment effect of the combination of A_1 and A_2 is given by $\theta_j(\boldsymbol{\beta}) = \beta_0 + \beta_1 + \beta_2 + \beta_{1,2}$; and the treatment effect of the three-agent combination of A_1, A_2 and A_3 is $\theta_j(\boldsymbol{\beta}) = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_{1,2} + \beta_{1,3} + \beta_{2,3} + \beta_{1,2,3}$. To be consistent with the lung cancer trial described in Section 1.2, we assume that a smaller value of $\theta_j(\boldsymbol{\beta})$ (i.e., less symptom burden) represents a better response. Then the hypotheses can be formally expressed as

$$H_j : \theta_j(\boldsymbol{\beta}) = \min(\theta_0(\boldsymbol{\beta}), \cdots, \theta_{p-1}(\boldsymbol{\beta})), \quad j = 0, \cdots, p-1.$$

We let $\pi_j(\boldsymbol{\beta}, \sigma^2)$ denote the prior distribution assigned to the unknown parameters $\boldsymbol{\beta}$ and σ^2 under H_j. Further discussion of the prior specification is provided in Section 2.3; for the moment we note that the domain of each prior is restricted to values of

$\boldsymbol{\beta}$ that are consistent with the hypothesis under which they are defined [21]. Given these prior densities, the marginal density of the observed data $\boldsymbol{y}$ under H_j is

$$m_j(\boldsymbol{y}) = \int\int f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2)\pi_j(\boldsymbol{\beta}, \sigma^2)\mathrm{d}\sigma^2\mathrm{d}\boldsymbol{\beta}, \tag{3.2}$$

and the Bayes factor [23, 15] of H_i to H_j is given by

$$B_{ij} = \frac{m_i(\boldsymbol{y})}{m_j(\boldsymbol{y})}. \tag{3.3}$$

If $p(H_j)$ denotes the prior probability of H_j, then the posterior probability of H_j given the data $\boldsymbol{y}$ is

$$p(H_j|\boldsymbol{y}) = \frac{p(H_j)m_j(\boldsymbol{y})}{\sum_{i=0}^{p-1} p(H_i)m_i(\boldsymbol{y})} = \left[\sum_{i=0}^{p-1} \frac{p(H_i)}{p(H_j)}B_{ij}\right]^{-1}. \tag{3.4}$$

If we assume that all hypotheses are equally likely *a priori*, then the posterior probability of H_j simplifies to

$$p(H_j|\boldsymbol{y}) = \frac{m_j(\boldsymbol{y})}{\sum_{i=0}^{p-1} m_i(\boldsymbol{y})} = \left[\sum_{i=0}^{p-1} B_{ij}\right]^{-1}. \tag{3.5}$$

The value of $p(H_j|\boldsymbol{y})$ has a very intuitive probability interpretation—the probability that the jth combination is the best treatment conditional on the observed data. Meanwhile, the value of $p(H_0|\boldsymbol{y})$ is the probability that the placebo is the best treatment. Therefore, it provides a natural evidence-based mechanism to adaptively assign patients to efficacious combinations and select the most promising combination.

3.1.2 Trial Design

We propose the following adaptive randomization scheme for the conduct of the trial. We assume that a total of N patients are available for testing, and that the first $m \times p$ patients are equally randomized into the p treatment arms using m

replications of a complete factorial design, i.e., m patients are randomized to each of p arms. The advantage of using a factorial design is that it allows us to rapidly obtain preliminary estimates of the main treatment effects. Following the lead-in factorial phase of the design, subsequent patients are assigned to a treatment according to the posterior probability that each treatment is best. The resulting design can be described as follows.

1. Assign $m \times p$ patients to the p treatment arms using m replications of a factorial design.
2. For $i = m \times p + 1, \ldots, N$, randomize the ith patient to the jth treatment arm with probability $p(H_j|\boldsymbol{y})$, $j = 0, \ldots, p-1$, where $\boldsymbol{y} = (y_1, \cdots, y_{i-1})'$ are the observed outcomes data from the first $i-1$ patients.
3. At the end of the trial, we select the combination j^* that has the highest posterior model probability, i.e., $j^* = \text{argmax}_j p(H_j|\boldsymbol{y}), j = 1, \cdots, p-1$.

During the trial, we impose the following futility stopping rule: the trial is terminated for futility if

$$\max\{p(\theta_0 - \theta_j > \delta|\boldsymbol{y})\} < \alpha, j = 1, \cdots, p-1$$

where δ and α are the prespecified minimal effect size and threshold, respectively. That is, at any time during the trial, conditional on the observed data, if the probability of achieving an effect size of δ for the best treatment arm is below the threshold α, we terminate the trial. In practice, the value of δ can be elicited from investigators, and the values of design parameters m and α can be chosen by examining the operating characteristics of the trial in simulation studies.

3.1.3 Delayed Outcomes

In general, outcome-dependent adaptive randomization, such as the one we have proposed, assumes that the outcome is quickly ascertainable so that when an incoming patient is ready for randomization, the previous patients have been assessed and their outcomes are completely observed. This assumption may not hold in practice. In many cases, the patient outcomes require a long follow-up time to be assessed (or the accrual is fast), so their outcomes are not available when a new patient is randomized. To address this delayed outcome issue, one approach is to suspend accrual and wait until the outcomes of patients treated in the trial are fully observed. However, this approach is often not practical because it causes lengthy delays in a trial, wastes patient resources, and causes administrative problems. Alternatively, we propose to base our adaptive randomization scheme only on those patient outcomes that are available at the time that each new patient is randomized. Our simulation studies in Section 3.2.1 show that, with finite samples, this observed-data approach is competitive to the approach of suspending accrual.

3.1.4 Prior Specification and Derivation of Bayes Factor

We adopt the encompassing prior approach proposed by Klugkist et al. [25] and Klugkist and Hoijtink [24] to set the prior distributions on $\boldsymbol{\beta}$ and σ^2 under each hypothesis. In this approach, we first specify a prior distribution for the unconstrained model, and then based on that prior define prior densities under each hypothesis. More specifically, we begin by assigning a noninformative prior to σ^2 of the form $\pi(\sigma^2) \propto 1/\sigma^2$. Given σ^2 and a hyperparameter g, we assume that $\boldsymbol{\beta}$ has a normal prior density of the form $\pi(\boldsymbol{\beta}|\sigma^2) \sim N(\mathbf{0}, g\sigma^2\boldsymbol{I}_p)$, where $\boldsymbol{I}_p$ denotes a $p \times p$ identity matrix.

To modify the unconstrained priors for application to hypothesis H_j, $j = 0, \ldots, p-1$, we restrict the domain of $\boldsymbol{\beta}$ under each hypothesis so that it is consistent with the assumptions of the given hypothesis [21]. That is, under hypothesis H_j, the domain of $\boldsymbol{\beta}$ is restricted to the value space satisfying the condition $\theta_j(\boldsymbol{\beta}) = \min(\theta_0(\boldsymbol{\beta}), \cdots, \theta_{p-1}(\boldsymbol{\beta}))$. This leads to the encompassing prior for H_j defined according to

$$\pi_j(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \pi_j(\boldsymbol{\beta}|\sigma^2) = \frac{1}{c_j} N(\mathbf{0}, g\sigma^2 \boldsymbol{I}_p) I_{\min}(\theta_j(\boldsymbol{\beta})), \tag{3.6}$$

where $I_{\min}(\theta_j(\boldsymbol{\beta}))$ denotes an indicator function of whether $\theta_j(\boldsymbol{\beta}) = \min(\theta_0(\boldsymbol{\beta}), \cdots, \theta_{p-1}(\boldsymbol{\beta}))$, and

$$c_j = \int N(\boldsymbol{\beta}|0, g\sigma^2 \boldsymbol{I}_p) I_{\min}(\theta_j(\boldsymbol{\beta})) \mathrm{d}\boldsymbol{\beta}.$$

The prior densities used to define each hypothesis are thus non-local with respect to one another, which enables us to more rapidly exclude hypotheses that are inconsistent with the data [21].

Under model (3.1) and the encompassing prior (3.6), the marginal density of data $\boldsymbol{y}$ under hypothesis H_j is given by

$$\begin{aligned} m_j &= \iint f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) \pi_j(\boldsymbol{\beta}|\sigma^2) \pi(\sigma^2) \mathrm{d}\sigma^2 \mathrm{d}\boldsymbol{\beta} \\ &= \iint N(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I_n}) \frac{1}{c_j} N(\boldsymbol{\beta}|\mathbf{0}, g\sigma^2 \boldsymbol{I}_p) I_{\min}(\theta_j(\boldsymbol{\beta})) \frac{1}{\sigma^2} \mathrm{d}\sigma^2 \mathrm{d}\boldsymbol{\beta} \end{aligned}$$

where

$$c_j = \int N(\boldsymbol{\beta}|0, g\sigma^2 \boldsymbol{I}_p) I_{\min}(\theta_j(\boldsymbol{\beta})) \mathrm{d}\boldsymbol{\beta}.$$

Letting $\boldsymbol{\beta'} = \boldsymbol{\beta}/\sigma$, it follows that the normalizing constant

$$c_j = \int N(\boldsymbol{\beta'}|\mathbf{0}, g\boldsymbol{I_p}) I_{\min}(\theta_j(\sigma\boldsymbol{\beta'})) \mathrm{d}\boldsymbol{\beta'}$$

Recall that $\theta_j(\boldsymbol{\beta}), j = 0, \cdots, p-1$, is a linear function of $\boldsymbol{\beta}$, thus $I_{\min}(\theta_j(\sigma\boldsymbol{\beta'})) = I_{\min}(\sigma\theta_j(\boldsymbol{\beta'}))$. Since $\sigma > 0$, the order of $\{\sigma\theta_j(\boldsymbol{\beta'})\}$ is the same as that of $\{\theta_j(\boldsymbol{\beta'})\}$.

Thus

$$I_{\min}(\theta_j(\sigma\boldsymbol{\beta}')) = I_{\min}(\theta_j(\boldsymbol{\beta}'))$$

and

$$c_j = \int N(\boldsymbol{\beta}'|\mathbf{0}, g\boldsymbol{I_p}) I_{\min}(\theta_j(\boldsymbol{\beta}'))\mathrm{d}\boldsymbol{\beta}'.$$

We can see that c_j is independent of σ^2 , which greatly simplifies the evaluation of the marginal density of $\boldsymbol{y}$. Then it follows that

$$\begin{aligned}
m_j &= \iint N(\boldsymbol{y}|\boldsymbol{X\beta}, \sigma^2\boldsymbol{I_n})\frac{1}{c_j}N(\boldsymbol{\beta}|\mathbf{0}, g\sigma^2\boldsymbol{I}_p) I_{\min}(\theta_j(\boldsymbol{\beta}))\frac{1}{\sigma^2}\mathrm{d}\sigma^2\mathrm{d}\boldsymbol{\beta} \\
&= \frac{1}{c_j}\iint N(\boldsymbol{y}|\boldsymbol{X\beta}, \sigma^2\boldsymbol{I_n})N(\boldsymbol{\beta}|\mathbf{0}, g\sigma^2\boldsymbol{I}_p)\frac{1}{\sigma^2} I_{\min}(\theta_j(\boldsymbol{\beta}))\mathrm{d}\sigma^2\mathrm{d}\boldsymbol{\beta} \\
&= \frac{\Gamma((n+p)/2)}{c_j g^{p/2}\pi^{(n+p)/2}}\int\left((\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) + \frac{1}{g}\boldsymbol{\beta}^T\boldsymbol{\beta}\right)^{-(n+p)/2} I_{\min}(\theta_j(\boldsymbol{\beta}))\mathrm{d}\boldsymbol{\beta} \\
&= \frac{\Gamma(n/2)\sqrt{|\boldsymbol{V}|}}{c_j g^{p/2}\pi^{n/2}(\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{\mu}^T\boldsymbol{V}^{-1}\boldsymbol{\mu})^{n/2}}\int t_\nu(\Sigma, \boldsymbol{\mu}) I_{\min}(\theta_j(\boldsymbol{\beta}))\mathrm{d}\boldsymbol{\beta}
\end{aligned}$$

Therefore the Bayes factor of H_i to H_j, B_{ij}, is given by

$$B_{ij} = \frac{m_i(\boldsymbol{y})}{m_j(\boldsymbol{y})} = \frac{r_i/c_i}{r_j/c_j}.$$

where

$$r_j = \int t_\nu(\Sigma, \boldsymbol{\mu}) I_{\min}(\theta_j(\boldsymbol{\beta}))\mathrm{d}\boldsymbol{\beta}.$$

Here $t_\nu(\cdot)$ denotes a multivariate student distribution with degree of freedom $\nu = n$, scale matrix $\Sigma = \boldsymbol{V}(\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{\mu}^T\boldsymbol{V}^{-1}\boldsymbol{\mu})/n$ and median $\boldsymbol{\mu} = \boldsymbol{V}\boldsymbol{X}^T\boldsymbol{y}$, where $\boldsymbol{V} = (\frac{1}{g}\boldsymbol{I}_p + \boldsymbol{X}^T\boldsymbol{X})^{-1}$, $\boldsymbol{X}$ is the design matrix in model (3.1) and n is the number of patients who have completed the assessment during the course of the trial.

3.2 Numerical Studies

3.2.1 Operating Characteristics

We evaluated the operating characteristics of the proposed Bayesian adaptive screening (BAS) trial design through extensive simulation studies. In the context of the lung cancer trial, we considered a total of 16 combinational treatments, including the placebo control, that result from 4 agents (Table 3.1). Two hundred patients were available for enrollment (i.e., N=200), and we performed $m = 2$ replications of the factorial design to obtain preliminary estimates of the treatment effects. The accrual rate was 12 patients per month, and it took 10 days to obtain the symptoms outcome. Because the accrual was fast, we faced the delayed-outcome problem, that is, when a new patient is accrued and ready for randomization, some patients treated in the trial may have not finished their 10-day assessment and their outcomes are not available for calculating the randomization probabilities for the new patient. To deal with this issue, we adopted the observed-data approach described previously and calculated the randomization probabilities based on observed data when the outcomes of some patients are not available. Because the observed-data approach supports continuous accrual, it took approximately 17 months to complete the trial. For this trial, the approach of suspending accrual apparently is not feasible because it would lead to an infeasibly long trial lasting at least 4.8 years. Although the accrual-suspension approach is not useful in practice, it provides a theoretical benchmark for comparison because it represents the optimal case that the complete data are available to determine treatment assignment. For convenience, we denote the BAS design based on the accrual-suspension approach as BAS_{susp}. We configured the simulation parameters, $\boldsymbol{\beta}$, to generate 12 different efficacy scenarios. The simulation results of the selection percentage of each treatment under these 12 different efficacy scenarios are

displayed in Table 3.2. The total selection percentage of target treatments and the total percentage of patients assigned to targets are summarized in Table 3.3. Under each scenario, the most efficacious combination was defined as the combination with the smallest value of symptom burden, $\theta(\boldsymbol{\beta})$. We set the residual variance $\sigma^2 = 130$ based on previous symptom report data. The two parameters involved in the futility stopping rule, α and δ, were set to 0.35 and 10, respectively. We also compared the proposed BAS design to a design based on 12 replications of the complete factorial design on the 16 treatments, randomly allocating the last 8 available patients to treatments. For the factorial design (FD), the treatment with the lowest value of the least square estimate of $\theta(\boldsymbol{\beta})$ was selected as the best treatment at the end of the trial. We carried out 2,000 simulations for each scenario.

Table 3.1: The 16 combinations of four agents (A_1, A_2, A_3 and A_4) investigated in the lung cancer trial.

Treatment	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7
A_1	0	1	0	0	0	1	1	1
A_2	0	0	1	0	0	1	0	0
A_3	0	0	0	1	0	0	1	0
A_4	0	0	0	0	1	0	0	1
Treatment	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}
A_1	0	0	0	1	1	1	0	1
A_2	1	1	0	1	1	0	1	1
A_3	1	0	1	1	0	1	1	1
A_4	0	1	1	0	1	1	1	1

Scenarios 1 to 4 simulated the cases in which there was a single best treatment. In scenario 1, the best (or most efficacious) treatment was T_1 (i.e., single agent A_1), and the BAS design substantially outperformed FD. The selection probability of the target treatment T_1 under the BAS design was 86.1%, while under FD it was 64.9%. In addition, compared to FD, the BAS design allocated a significantly higher percentage of patients to the best treatment (6.3% versus 39.3%, respectively). The performance of the BAS design was rather similar to that of the optimal BAS_{susp} design. The selection probability of the target treatment under the BAS design was only 1.1% lower than that of BAS_{susp} design, and the percentage of patients allocated to the best treatment were almost same in two designs (39.3% versus 39.1%, respectively), suggesting that randomization based on observed data provided an efficient way to handle delayed outcomes. In scenario 2, the best treatment was T_5, the combination of agents A_1 and A_2. In this case, the selection probability of the BAS design was 21.5% higher than that of FD, and the BAS design assigned 32.8% more patients to the best treatment. In scenarios 3 and 4, the three-drug combination T_{11} (i.e., combination of A_1, A_2 and A_3) and the four-drug combination T_{15} were defined as the optimal treatments. Comparisons under these scenarios were similar to those made under scenarios 1 and 2. The selection probabilities of the BAS design were more than 18% higher than these of FD, and the percentages of patients assigned to the best treatment under the BAS design were more than 33% higher than those under FD. Again, we observed that the performance of the BAS design was rather similar to that of the BAS_{susp} design.

Scenarios 5 to 8 were designed to evaluate the performance of the design when there were two best treatments that were equally effective. In scenario 5, T_1 and T_2 were the target treatments with the highest efficacy. The BAS design selected T_1 with

a probability of 46.7% and T_2 with a probability of 47.5%; whereas FD selected these two targets with probabilities of 41.0% and 43.1%, respectively. That is, the total selection probability of T_1 and T_2 under the BAS design was 10.1% higher than that under FD. The percentage of patients assigned to the best treatment using the BAS design was 38.8% higher than that using FD. For scenarios 6 to 8, the BAS design again outperformed FD, achieving substantially higher selection probabilities and assigning higher percentages of patients to the optimal treatments. Scenarios 9 and 10 had three optimal treatments, and scenario 11 had four target treatments. Under these scenarios, the performance of the proposed BAS design once again dominated that of FD. Compared to FD, the total selection probabilities of the target treatments under the BAS design were improved by 0.7-5.9%, and the percentages of patients assigned to the best treatments were improved by 39.8-44.6%. Scenario 12 represents the case in which the treatment effects of all combinations are the same as that of the placebo. Under this scenario, the BAS design and FD terminated the trial due to futility, with respective probabilities of 85.2% and 90.5%.

As demonstrated in the simulation study, the proposed BAS design achieved two important clinical goals simultaneously. First, it selected the best treatment arms with high probability. Second, it allocated more patients to the best treatments. This result seems somewhat surprising because the common notion is that these two goals are in conflict with each other. That is, it is often assumed that randomization schemes in which patients are allocated to effective treatments have less power to detect the best treatment at the end of the trial. This may be the case in comparisons of only two or three treatments, but in more complicated settings in which large numbers of treatments and treatment combinations are tested, the BAS offers significant gains in both power and patient allocation.

The success of our adaptive randomization scheme in accomplishing both goals simultaneously in high-dimension settings can be understood by noting that our design allocates more patients to the subset of treatments that are competitive. By reallocating patients away from ineffective treatments, we obtain higher power to distinguish between the top treatments. For example in scenario 1, our design allocated 39.3 and 10.3 patients to the best and second best treatments (T_1 and T_{15}); in contrast, FD allocated 6.3 patients to both T_1 and T_{15}. As a consequence, BAS had higher power to distinguish between T_1 and T_{15}.

Table 3.2: Operating characteristics of the proposed Bayesian adaptive screening (BAS) design, BAS design based on the accrual-suspension approach (BAS_{susp}) and factorial design (FD). The efficacious treatments are bolded.

Treatment Effect				Selection percentage of each treatment											
θ_0	θ_1	θ_2	θ_3												
θ_4	θ_5	θ_6	θ_7												
θ_8	θ_9	θ_{10}	θ_{11}												
θ_{12}	θ_{13}	θ_{14}	θ_{15}	BAS				BAS_{susp}				FD			
				Scenario 1											
0	**-25**	-13	-12	0.0	**86.1**	0.0	0.0	0.0	**87.2**	0.0	0.0	0.0	**64.9**	0.0	0.0
-11	-20	-19	-18	0.0	2.5	1.4	0.5	0.0	2.4	1.0	0.2	0.0	8.9	4.4	2.6
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0
-16	-17	-14	-21	0.2	0.2	0.1	6.2	0.1	0.2	0.0	6.0	0.2	1.6	0.3	16.1
				Scenario 2											
0	-20	-12	-13	0.0	2.8	0.0	0.0	0.0	2.8	0.0	0.1	0.0	10.0	0.0	0.0
-11	**-25**	-19	-18	0.0	**86.6**	1.4	0.5	0.0	**85.4**	1.5	0.5	0.0	**65.1**	4.2	3.1
-14	-15	-12	-10	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.5	0.0	0.0
-16	-17	-14	-21	0.2	0.3	0.0	5.5	0.2	0.2	0.0	6.6	0.4	1.6	0.2	14.2
				Scenario 3											
0	-10	-12	-13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
-11	-20	-19	-18	0.0	3.3	1.2	0.6	0.0	2.5	1.1	0.7	0.0	8.6	4.5	2.9
-14	-15	-12	**-25**	0.0	0.0	0.0	**86.0**	0.0	0.0	0.0	**87.5**	0.1	0.2	0.1	**67.6**
-14	-17	-14	-21	0.1	0.1	0.2	6.2	0.0	0.2	0.0	5.7	0.2	1.3	0.1	13.5
				Scenario 4											
0	-16	-12	-13	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.7	0.0	0.0
-11	-20	-19	-18	0.0	3.5	1.0	0.2	0.0	2.1	1.0	0.5	0.0	9.1	4.1	2.7
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.4	0.0	0.0
-16	-21	-14	**-25**	0.0	5.7	0.0	**86.8**	0.1	6.5	0.0	**86.5**	1.0	14.4	0.1	**66.6**
				Scenario 5											
0	**-25**	**-25**	-13	0.0	**46.7**	**47.5**	0.0	0.0	**46.1**	**48.8**	0.0	0.0	**41.0**	**43.1**	0.0
-11	-20	-19	-18	0.0	1.0	0.2	0.2	0.0	0.6	0.3	0.0	0.0	4.1	2.0	1.4
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-16	-17	-14	-21	0.0	0.1	0.0	2.5	0.0	0.1	0.0	2.3	0.4	0.5	0.1	6.8
				Scenario 6											
0	**-25**	-12	-13	0.0	**46.2**	0.0	0.0	0.0	**47.0**	0.0	0.0	0.0	**43.9**	0.0	0.0
-11	**-25**	-19	-18	0.0	**49.1**	0.2	0.1	0.0	**48.8**	0.2	0.2	0.0	**43.0**	1.8	1.1
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-16	-17	-14	-21	0.1	0.0	0.0	1.7	0.0	0.0	0.0	1.5	0.2	0.9	0.0	8.3

Table 3.2 continued.

Treatment Effect				Selection percentage of each treatment											
θ_0	θ_1	θ_2	θ_3												
θ_4	θ_5	θ_6	θ_7												
θ_8	θ_9	θ_{10}	θ_{11}												
θ_{12}	θ_{13}	θ_{14}	θ_{15}	BAS				BAS_{susp}				FD			
				Scenario 7											
0	**-25**	-12	-13	0.0	**46.3**	0.0	0.0	0.0	**47.9**	0.0	0.0	0.0	**42.1**	0.0	0.0
-11	-20	-19	-18	0.0	0.8	0.4	0.2	0.0	0.8	0.2	0.2	0.0	4.3	2.1	0.9
-14	-15	-12	**-25**	0.0	0.0	0.0	**48.5**	0.0	0.0	0.0	**46.8**	0.0	0.0	0.0	**41.3**
-16	-17	-14	-21	0.0	0.0	0.0	1.8	0.0	0.2	0.0	2.0	0.2	0.7	0.0	7.7
				Scenario 8											
0	**-25**	-12	-13	0.0	**47.5**	0.0	0.0	0.0	**49.4**	0.0	0.0	0.0	**43.8**	0.0	0.2
-11	-20	-19	-18	0.0	0.8	0.5	0.1	0.0	0.4	0.0	0.2	0.0	5.1	2.5	1.8
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0
-16	-17	-14	**-25**	0.1	0.1	0.0	**48.7**	0.0	0.0	0.0	**48.3**	0.2	0.9	0.1	**44.5**
				Scenario 9											
0	**-25**	**-25**	-13	0.0	**31.3**	**31.5**	0.0	0.0	**33.4**	**30.9**	0.0	0.0	**31.1**	**32.0**	0.0
-11	**-25**	-19	-18	0.0	**34.1**	0.3	0.0	0.0	**33.4**	0.2	0.0	0.0	**29.2**	1.2	0.5
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-16	-17	-14	-21	0.1	0.0	0.0	1.0	0.0	0.0	0.0	0.9	0.2	0.4	0.0	4.7
				Scenario 10											
0	**-25**	**-25**	**-25**	0.0	**30.9**	**33.7**	**32.2**	0.0	**30.6**	**34.1**	**31.3**	0.0	**30.4**	**30.9**	**29.6**
-11	-20	-19	-18	0.0	0.4	0.5	0.2	0.0	0.6	0.5	0.2	0.0	2.5	1.2	0.4
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-16	-17	-14	-21	0.0	0.0	0.0	1.1	0.0	0.0	0.0	1.1	0.1	0.3	0.0	4.2
				Scenario 11											
0	**-25**	-12	-13	0.0	**23.5**	0.0	0.0	0.0	**26.3**	0.0	0.0	0.0	**25.9**	0.0	0.0
-11	**-25**	-19	-18	0.0	**25.7**	0.2	0.2	0.0	**25.4**	0.2	0.0	0.0	**22.7**	0.8	0.6
-14	-15	-12	**-25**	0.0	0.0	0.0	**24.7**	0.0	0.0	0.0	**24.6**	0.0	0.0	0.0	**25.2**
-16	-17	-14	**-25**	0.0	0.0	0.0	**24.6**	0.0	0.0	0.1	**22.7**	0.0	0.2	0.0	**24.0**
				Scenario 12											
0	0	0	0	0.0	0.9	1.1	0.6	0.0	0.8	1.1	0.9	0.0	0.8	0.5	0.4
0	0	0	0	1.0	0.8	1.5	1.4	1.0	0.8	0.8	1.0	0.4	1.0	0.4	0.8
0	0	0	0	1.0	0.8	0.8	1.1	1.0	1.0	1.2	0.8	0.8	0.5	1.0	0.6
0	0	0	0	0.8	1.1	0.8	1.1	1.1	1.1	1.2	1.1	0.7	0.4	0.4	0.8

Table 3.3: Summary of the simulation results, including the total selection percentage of target treatments and percentage of patients treated at the target treatments.

Scenario	1	2	3	4	5	6	7	8	9	10	11	12
						Total selection percentage of target treatments						
BAS	86.1	86.6	86.0	86.8	94.2	95.3	94.8	96.2	96.9	96.8	98.5	85.2*
BAS_{susp}	87.2	85.4	87.5	86.5	94.9	95.8	94.7	97.7	97.7	96.0	99.0	85.1*
FD	64.9	65.1	67.6	66.6	84.1	86.9	83.4	88.3	92.3	90.9	97.8	90.5*
						Total percentage of patients treated at target treatments						
BAS	39.3	39.0	39.2	39.8	51.3	54.7	51.2	55.9	61.5	58.6	69.6	
BAS_{susp}	39.1	38.7	40.4	40.0	51.8	54.6	51.2	56.8	61.6	58.5	69.5	
FD	6.3	6.2	6.2	6.3	12.5	12.5	12.5	12.5	18.8	18.8	25.0	

* the percentage of trials terminated due to futility.

3.2.2 Sensitivity Analysis

The encompassing prior for $\boldsymbol{\beta}$ requires pre-specification of the value for hyperparameter g. We conducted a sensitivity analysis to check the robustness of the design to the value of g. Specifically, we considered a tighter (or more informative) prior with $g = 5$ and a more diffused (or noninformative) prior with $g = 20$. Table 3.4 shows the results under scenarios 2, 4, 6, 8 and 10. Under each of these scenarios, the results with $g = 5$ or 20 were very similar to these reported in Table 3.2 (with $g = 10$), suggesting that the operating characteristics of the proposed design were not sensitive to the specification of g as long as it was reasonably diffuse. For example, in scenario 2, the selection probabilities of the target treatment, T_1, were 87.3% and 87.1% under $g = 5$ and 20, respectively, which were very similar to that under $g = 10$ (86.6%). The percentages of patients assigned to T_1 were also very similar for $g = 5, 10$ and 20.

We conducted another sensitivity analysis to examine the performance of the proposed design when the outcome needs a longer assessment period to be evaluated. We assumed an assessment period of 60 days and an accrual rate of 6 patients per

month. As shown in Table 3.5, the results were very similar to these reported in Table 3.2, in which the assessment period was 10 days with an accrual rate of 12 patients per month. This suggests that the proposed design is robust to the length of the assessment period and delayed outcomes.

Table 3.4: Sensitivity analysis with different values of g under scenarios 2, 4, 6, 8 and 10 for the proposed Bayesian adaptive screening (BAS) design. The efficacious treatments are bolded.

	$g = 5$				$g = 20$			
Scenario	Selection percentage				Selection percentage			
2	0.0	2.6	0.0	0.0	0.0	2.8	0.0	0.0
	0.0	**87.3**	1.1	0.5	0.0	**87.1**	1.2	0.4
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.1	0.2	0.0	5.9	0.2	0.3	0.2	5.1
4	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0
	0.0	2.5	0.7	0.2	0.0	2.9	0.8	0.6
	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
	0.2	5.0	0.0	**88.2**	0.2	5.9	0.0	**86.2**
6	0.0	**48.4**	0.0	0.0	0.0	**45.1**	0.0	0.0
	0.0	**46.5**	0.4	0.2	0.0	**50.5**	0.3	0.2
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.1	0.0	1.8	0.0	0.0	0.0	1.7
8	0.0	**49.5**	0.0	0.0	0.0	**48.6**	0.0	0.0
	0.0	0.8	0.4	0.2	0.0	0.6	0.3	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.2	0.0	**47.0**	0.0	0.0	0.0	**48.6**
10	0.0	**32.2**	**32.1**	**32.1**	0.0	**32.9**	**32.1**	**31.3**
	0.0	0.4	0.4	0.2	0.0	0.5	0.4	0.1
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	1.2	0.0	0.0	0.0	1.1

Table 3.5: Sensitivity analysis for the proposed Bayesian adaptive screening (BAS) design with an assessment period of 60 days and an accrual rate of 6 patients per month. The efficacious treatments are bolded.

Selection percentage				Selection percentage				Selection percentage				Selection percentage			
Scenario 1				Scenario 2				Scenario 3				Scenario 4			
0.0	**86.5**	0.0	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
0.0	3.2	1.1	0.5	0.0	**85.3**	1.4	0.6	0.0	2.7	1.3	0.2	0.0	2.8	1.1	0.9
0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	**86.0**	0.0	0.0	0.0	0.0
0.2	0.3	0.0	5.3	0.1	0.3	0.2	5.7	0.0	0.2	0.2	6.6	0.2	6.2	0.0	**84.8**
Scenario 5				Scenario 6				Scenario 7				Scenario 8			
0.0	**48.5**	**45.6**	0.0	0.0	**47.9**	0.0	0.0	0.0	**47.2**	0.0	0.0	0.0	**48.4**	0.0	0.0
0.0	0.7	0.3	0.4	0.0	**47.0**	0.3	0.4	0.0	0.8	0.6	0.4	0.0	0.8	0.2	0.2
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**46.2**	0.0	0.0	0.0	0.0
0.0	0.2	0.0	1.7	0.0	0.0	0.0	1.6	0.0	0.0	0.0	2.6	0.1	0.1	0.0	**47.8**
Scenario 9				Scenario 10				Scenario 11				Scenario 12			
0.0	**31.9**	**31.9**	0.0	0.0	**32.3**	**32.0**	**31.8**	0.0	**24.2**	0.0	0.1	0.0	1.0	1.0	1.2
0.0	**32.4**	0.2	0.2	0.0	0.6	0.4	0.1	0.0	**24.6**	0.4	0.2	0.9	1.1	1.2	1.1
0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	**25.0**	1.2	1.0	1.1	1.1
0.0	0.2	0.0	1.3	0.0	0.0	0.0	1.4	0.0	0.1	0.0	**24.5**	0.9	1.0	1.1	1.1

CHAPTER 4

A Bayesian Phase II Design with Continuous Monitoring for Late-Onset Responses Using Multiple Imputation

In this chapter, we propose a Bayesian single-arm phase II design with continuous monitoring for late-onset responses. The interim monitoring rule is employed to terminate the trial early for futility if there is sufficient evidence to determine the inefficiency of experimental agents. The benefits of the interim monitoring include avoiding assigning an unacceptable number of patients to inferior treatments and saving resources. In general, interim monitoring based on previous responses assumes that the outcome could be observed shortly after the initiation of treatment. Therefore, at the decision-making time, the outcomes of previous enrolled patients have been completely observed. However, this assumption may not hold. For late-onset responses, patient outcomes may occur long after the assignment of treatment. With fast accrual rate, the amount of missing responses at the decision-making time is large.

To address the issue of late-onset responses, we propose an approach built on missing data methodology to handle the missing responses and apply standard methods to estimate the response rate. Specifically, we use a piece-wise exponential model to estimate the hazard function of time to response data and use the multiple imputation method to deal with unobserved responses. For the proposed methods,

we do not need to interrupt patient accrual to wait for the full observation of previously patients that dramatically shortens the trial duration. We conducted extensive simulation studies to evaluate the operating characteristics of the proposed method. The comparison with standard, observed and complete methods show that the proposed method reduces the total length of the trial duration and yields a desirable operating characteristics for different physician-specified lower bounds of response rate with different true response rates.

In following sections, we introduce the probability models and trial design with interim monitoring. We propose a multiple imputation method to handle the missing responses. We examine the operating characteristics of our proposed design through extensive simulation studies and sensitivity analyses.

4.1 Methods

4.1.1 Probability model

Considering a single-arm phase II trial, we assume that subjects enter the trial sequentially and each subject will be assessed for a fixed assessment period of T. We consider a binary response as a primary outcome variable for subject i during the follow-up time, denoted by y_i, where $y_i = 1$ if treatment-related response is observed in $(0, T)$ and $y_i = 0$ otherwise. The length of assessment time T is chosen based on previous knowledge to ensure that a treatment-related response event usually occur within $(0, T)$. For different diseases and treatment agents, the evaluation period T varies from days to months.

During the stage of the trial, suppose that n patients have entered the trial, and let y_i denote the binary response outcome for ith subject. Denoting the observed response data for n subjects by $\boldsymbol{y} = \{y_i, i = 1, \cdots, n\}$, the likelihood function is given

by

$$L(\boldsymbol{y}|\pi) = \prod_{i=1}^{n} \pi^{y_i}\{1-\pi\}^{1-y_i}, \tag{4.1}$$

where π denotes the response rate of the experimental treatment. Letting $f(\pi)$ denotes a prior distribution for π, the posterior distribution of π is given by

$$f(\pi|\boldsymbol{y}) = \frac{L(\boldsymbol{y}|\pi)f(\pi)}{\int L(\boldsymbol{y}|\pi)f(\pi)\mathrm{d}\pi}. \tag{4.2}$$

For the conjugate prior specification of response rate π, we set $f(\pi)$ as a beta distribution with two shape parameters ζ and ξ, then the posterior distribution of π is a beta distribution with shape parameters $\zeta + \sum_{i=1}^{n} y_i$ and $\xi + n - \sum_{i=1}^{n} y_i$. Here, we suggest to use a vague or non-informative prior for π.

4.1.2 Interim Monitoring and Late-onset Responses

Interim monitoring is usually conducted to stop the trial early for futility if there is sufficient evidence to demonstrate the inefficiency of experimental drug. The monitoring rules can be applied to the trial continuously or after a group with a fixed number of subjects. The advantage of interim monitoring is that if the experimental treatment is deemed inefficacious, we can stop the trial earlier and assign fewer patients to the ineffective treatment.

The stopping rule in our trial design is based on a physician-specified lower bound of response rate for the experimental treatment, denoting by ϕ. If the true response rate of the experimental treatment is higher than the lower bound ϕ, we consider the experimental treatment to be efficacious and should continue recruiting new arriving patients into the trial; otherwise, the trial should be terminated when sufficient information has been collected to demonstrate its futility. In our trial design, we conduct continuous monitoring before each new patient entering the trial

after n_0 patients are completely followed. During the course of the trial, suppose that n ($n > n_0$) patients have entered the trial. If the posterior probability of response rate achieving the physician-specified lower bound, ϕ, is less than a cut-off value, ψ, e.g., $Pr(\pi < \phi) > \psi$, we stop the trial due to futility; otherwise the trial continues recruiting new patients until the exhaustion of the total sample size N and concludes that the experimental agent is sufficiently promising for further study. The inference of the posterior estimates of response rate π is made given observed patient data by the formula (4.2).

As mentioned before, continuous monitoring based on previous patients outcomes needs outcomes to be assessed quickly after the initiation of the treatment. However, it may not be the case for late-onset responses which may occur long after the assignment of treatment. Before we discuss our method to address this issue, we introduce the missing mechanism of the late-onset responses. In general, for late-onset responses, the assessment time T usually is longer than the interarrival time between two consecutive cohorts. Here, the interarrival time is defined as the interval time between the entering time of two consecutive cohorts. If we denote the patient interarrival time by τ, it indicates that when $\tau < T$, some patients under the treatment might have not yet exhibited responses or completed evaluation period when new patient is ready to enter the trial. Specifically, we denote the time to response by t_i for the ith subject and let u_i ($0 \leq u_i \leq T$) denote its actual follow-up time at the moment of interim monitoring. If the actual follow-up time is less than the true response time, i.e., $u_i < t_i$, it indicates that the patient response could not be observed at the moment of interim monitoring. Therefore, responses are missing only when patients have not yet experienced response ($u_i < t_i$) and have not fully followed up to T ($u_i < T$). If patients either have experienced responses ($t_i \leq u_i$)

or have completed followed-up ($u_i = T$) without experiencing responses, patients outcomes are observed. Yuan and Yin (2012) showed that under this missing data mechanism, the generated data are non-ignorable missing which means the probability of missingness of responses depends on the underlying missing outcomes. For patients who will not experience responses during the whole assessment period, they are more likely to have missing outcomes at the interim monitoring time compared to patients who experienced responses.

As we know if the patient accrual rate increases, there would be more missing responses at the same decision-making time. For example, considering a trial with an assessment period of 3 months, if the accrual rate is 2 patients per month, i.e., the interarrival time $\tau = 1/2$ months, there would be at most 6 missing outcomes at the decision-making time. If we increase the accrual rate to 4 patients per month, i.e., the interarrival time $\tau = 1/4$ months, there would be at most 12 missing outcomes at the decision-making time. Therefore, during the trial of conduct, the amount of missing data depends on the ratio of of the assessment period T and the interarrival time τ. We denote this ratio by A/I ratio $= T/\tau$. The larger the value of the ratio is, the greater the amount of missing data would be.

Comparing with missing completely at random or missing at random, non-ignorable missing data bring a new challenge to the trial design. To address the issue, one possible approach is to suspend the accrual and wait until the previously enrolled patients are fully followed-up. Therefore, the outcomes of all treated patients can be observed and there is no unobserved response before new patient entering the trial. Obviously, this method fully utilizes all available information and provides a precise estimate of response rate at the time of decision-making time. However, frequently suspension leads to an infeasibly long trial and brings inconvenience for

trial administration.

If we do not suspend the accrual and assign a newly arriving patient to experimental treatment immediately, those patients under treatment might not have completed the assessment period. A simple choice is to make early stopping decisions solely based on the outcomes from patients who have completed the assessment period or experienced responses during the assessment period. It means only the complete data thus far are used at the moment of interim monitoring and the data from the patients who have not completed the assessment period and not yet given responses are ignored. However, this method has a higher chance to include patients who would experience responses during the assessment period. The reason is that for patients who would not experience responses in $(0, T)$, their responses are more likely missing at the interim monitoring time compared to patients who would experience responses in $(0, T)$. Therefore, this method is problematic and overestimates the response rate. Another approach which does not suspend the accrual is to make inference based on data from all treated subjects. For patients who have not completed the assessment period, if there are no responses at the time of interim monitoring, the current outcomes of no responses will be considered as the final outcomes at the end of assessment period. Specifically, if the ith subject has not completed the evaluation period T and also has not experienced response, his/her response at the end of assessment period is considered as censored, i.e., $y_i = 0$. Although this approach includes all the patients under the treatment, it uses current observed information to replace the final outcomes for the partially observed patients. Due to the property of late-onset responses, these patients who have not completed the assessment period are more likely to give responses at the remaining assessment period. The longer we follow the patient, the higher probability that the patient will experience response

later. Therefore, this approach is also problematic since it often underestimates the response rate and terminates the trial unappropriate.

4.1.3 Multiple Imputation Approach

To address the missing-response issue introduced in previous section, we propose a method to handle the unobserved patient outcomes based on missing data theory [30]. Different from above simple methods, this method is built on missing data methodology and systematically treats the unobserved outcomes as missing data. Intuitively, we first fill in the missing data by the multiple imputation method and then apply the standard complete-data method to the imputed dataset. Multiple imputation provides a systematic way to impute the missing response data and meanwhile account for the sampling uncertainty due to the missing values [40]. Following this route, we replace each missing value with M imputed values, respectively, i.e., we impute the missing data M times to form M filled-in datasets. Then the standard complete-data methods can be applied to each of the filled-in datasets. By combining the M complete-data inferences, we take into account the imputation uncertainty.

To achieve the goal above, we specify a flexible piecewise exponential model for the time to response data during assessment period. Specifically, we consider a partition of the follow-up period $[0, T]$ into a finite number of K disjoint intervals $[0, h_1), [h_1, h_2), \cdots, [h_{K-1}, h_K = T]$ and assume a constant hazard λ_k in the kth interval. We define the observed time $x_i = \min(u_i, t_i)$ and $\delta_{ik} = 1$ if the ith subject experiences response in the kth interval; and $\delta_{ik} = 0$ otherwise. Let $\boldsymbol{\lambda} = \{\lambda_1, \cdots, \lambda_K\}$; when $\{x_i\}$ are completely observed, the likelihood function for n subjects based on

observed time to response data $\mathcal{D} = \{x_i, \delta_{ik}, i = 1, \cdots, n, k = 1, \cdots, K\}$ is given by

$$L(\mathcal{D}|\boldsymbol{\lambda}) = \prod_{i=1}^{n} \prod_{k=1}^{K} (\lambda_k)^{\delta_{ik}} exp\{-\lambda_k e_{ik}\}, \tag{4.3}$$

where $e_{ik} = h_k - h_{k-1}$ if $x_i > h_k$; $e_{ik} = x_i - h_{k-1}$ if $x_i \in [h_{k-1}, h_k)$; and otherwise $e_{ik} = 0$. K is the number of intervals defined for the piecewise exponential model. Large K results in a nonparametric model of hazard function and unstable estimates; small K lead to inadequate model fitting. In our simulation studies, we conducted sensitivity analysis with different values of K to check its robustness.

Let $f(\boldsymbol{\lambda})$ denote the joint prior distribution for all λ_i's. We write the joint posterior distribution as

$$f(\boldsymbol{\lambda}|\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda})}{\int L(\mathcal{D}|\boldsymbol{\lambda})}.$$

For the prior specification of the piece-wise exponential model, we adopt a correlated prior approach introduced in Qiou, Ravishanker and Dey (1999) [37]. Specifically, a discrete-time martingale process [1][2] is assigned to correlate the λ_i's in adjacent intervals, which introduces some smoothness to the estimates. Given $(\lambda_1, \cdots, \lambda_{k-1})$, we specify that

$$\lambda_k|\lambda_1, \cdots, \lambda_{k-1} \sim Gamma(c_k, \frac{c_k}{\lambda_{k-1}}), k = 1, \cdots, K$$

where $Gamma(\xi, \eta)$ represents a gamma distribution with a shape parameter ξ and a scale parameter η, so that $E(\lambda_k|\lambda_1, \cdots, \lambda_{k-1}) = \lambda_{k-1}$. The choice of the value of λ_0 is suggested as follows. We assume a constant hazard function for the whole assessment period, i.e., $\lambda_K = \lambda_{K-1} = \cdots = \lambda_0$. Then the value of λ_0 can be obtained by setting the response rate at the end of assessment period as the physician specified lower bound ϕ. The value of c_k indicates the amount of information for smoothness of λ_k. If $c_k = 0$, λ_k is independent of λ_{k-1} while if $c_k \to \infty$, $\lambda_k = \lambda_{k-1}$.

The conditional posterior distribution of λ_k given $\lambda_1, \cdots, \lambda_{k-1}$ and observed data $\mathcal{D}$ is

$$f(\lambda_k|\lambda_1, \cdots, \lambda_{k-1}, \mathcal{D}) = \frac{L(\lambda|\mathcal{D})f(\lambda_k|\lambda_1, \cdots, \lambda_{k-1})}{\int L(\lambda|\mathcal{D})f(\lambda_k|\lambda_1, \cdots, \lambda_{k-1})}.$$

After we obtained the full conditional distributions for each λ, the Gibbs sampler will be used to obtain posterior draws of unknown parameters for statistical inferences.

We denote the binary response outcomes for subjects in the trial by $\boldsymbol{y} = (\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis})$, where $\boldsymbol{y}_{obs}$ and $\boldsymbol{y}_{mis}$ denote the observed and missing response data, respectively. To carry out the multiple imputation, we draw the missing binary responses from its posterior predictive distribution which is given by

$$f(y_i|\mathcal{D}) = \int f(y_i|\boldsymbol{\lambda})f(\boldsymbol{\lambda}|\mathcal{D})\mathrm{d}\boldsymbol{\lambda}$$

The inference for informative missing responses is based on current observed time to response data $\mathcal{D}$ which is more informative than observed binary data $\boldsymbol{y}_{obs}$. The inference using only $\boldsymbol{y}_{obs}$ would lead to biased estimates. Specifically, we can draw M independent sampling of the unknown parameter $\boldsymbol{\lambda}$ with respect to its posterior distribution $f(\boldsymbol{\lambda}|\mathcal{D})$ given observed data $\mathcal{D}$. Therefore, based on the posterior estimates of $\boldsymbol{\lambda}$, we can calculate the response rate of the experimental treatment at any time t for $0 < t < T$. Generally, if the ith subject has not experienced response at the decision-making time given the actual follow-up time u_i, i.e., $t > u_i$, the conditional probability that the ith subject would experience response during the assessment period $(0, T)$ based on the posterior estimate of $\boldsymbol{\lambda}$ is

$$\hat{\pi}_i(\boldsymbol{\lambda}) = Pr(t < T|t > u_i) = \frac{Pr(t < T) - Pr(t < u_i)}{Pr(t > u_i)} = \frac{F(T) - F(u_i)}{F(u_i)}$$

where $\hat{\pi}_i(\boldsymbol{\lambda})$ denotes the conditional response rate and $F(\cdot)$ is the cumulative distribution function of the random variable. In our piece-wise exponential model, the

estimation of $F(s)$ for any $0 < s < T$ can be obtained by $F(s) = 1 - \sum_{k=1}^{K} exp(-\lambda_k e_k)$ where $e_k = h_k - h_{k-1}$ if $s > h_k$; $e_k = s - h_{k-1}$ if $s \in [h_{k-1}, h_k)$; and otherwise $e_k = 0$.

After we obtain the conditional response rate at the end of assessment period for subject i, $\hat{\pi}_i(\boldsymbol{\lambda})$, we can easily find that the full conditional distribution of binary response $y_i \in \boldsymbol{y}_{mis}$ is given by

$$f(y_i|\boldsymbol{\lambda}) = Bernoulli(\hat{\pi}_i(\boldsymbol{\lambda})),$$

where Bernoulli$(\cdot)$ represents the probability density function of a Bernoulli distribution. Based on M independent posterior samplings of the parameter $\boldsymbol{\lambda}$, we can draw M independent samplings of $y_i \in \boldsymbol{y}_{mis}$ from the above posterior distribution $f(y_i|\boldsymbol{\lambda})$. Here, the missing value y_i generated in this way is drawn from its posterior predictive distribution $f(y_i|\mathcal{D})$. We construct M imputed datasets by filling in $\boldsymbol{y}_{mis}$ with M independent samples $\boldsymbol{y}_{mis}^{(m)}, m = 1, \cdots, M$. Based on the mth imputed dataset $\boldsymbol{y}^{(m)} = \{\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis}^{(m)}\}$, we obtain the posterior distribution of the response rate $\pi^{(m)}$ by applying the simple Beta-Binomial model (4.1) and get the estimate of $Pr(\pi^{(m)} < \phi)$. Then we combine the estimates of the response rate across M imputed datasets by average and get the estimate of stopping criteria

$$Pr(\pi < \phi) = \frac{1}{M} \sum_{m=1}^{M} Pr(\pi^{(m)} < \phi).$$

If $Pr(\pi < \phi) > \psi$, we stop the trial due to futility; otherwise, the trial continues until the maximum sample size N is reached.

4.2 Numerical Studies

4.2.1 Operating Characteristics

In this chapter, we propose a multiple imputation method to handle unobserved responses at the decision-making time for a single-arm phase II trial design.

We conducted extensive simulation studies to evaluate the operating characteristics of the proposed method. A maximum number of $N = 50$ patients were treated sequentially. The assessment period was set as $T = 6$ months and the interarrival time between every two consecutive cohorts was $\tau = 1$ month, i.e., the A/I ratio=6. The number of intervals for piece-wise exponential model used in multiple imputation approach is assumed as $K = 6$. Specifically, we partitioned the assessment period $[0, T]$ into K equal intervals, i.e., $[0, \frac{1}{K}T), [\frac{1}{K}T, \frac{2}{K}T), \cdots, [\frac{K-1}{K}T, T]$. We assigned a beta distribution with $\zeta = 0.1$ and $\xi = 0.2$ for $f(\pi)$. Under each scenario, we simulated 1,000 trials.

We considered three different lower bounds of response rate with $\phi = 0.3, 0.4, 0.5$, respectively. Under each case of different lower bounds, we considered several scenarios with various true response rates of experimental treatment, denoted as $F_t(T)$. Specifically, for $\phi = 0.3$, we considered 5 scenarios with true response rates at 0.1, 0.2, 0.3, 0.4, and 0.5 respectively. For $\phi = 0.4$, we considered 5 scenarios with true response rates at 0.2, 0.3, 0.4, 0.5, and 0.6 respectively. Similarly, for $\phi = 0.5$, we considered 5 scenarios with true response rate as 0.3, 0.4, 0.5, 0.6, and 0.8 respectively.

Under each scenario, we generated time to response data from a Weibull distribution. To generate the late-onset responses with different degrees of responses occurring in latter half of the assessment period, $(T/2, T)$, we specify the Weibull distribution with different shape and scale parameters. Specifically, we choose the scale and shape parameters of the Weibull distribution based on the following two requirements. First, the true response rate at the end of follow-up indicates the value of cumulative distribution function at $t = T$, where t is generated from Weibull distribution. Second, the probability of occurring responses during $(T/2, T)$ is fixed

at a pre-specified percentage, which represents the degrees of late-onset responses. Therefore, based on the true response rate of experimental treatment and degrees of late-onset responses, we specify the shape and scale parameters of the Weibull distribution to generate the time to response data under each scenario. In our simulations, we assumed approximately 50%, 70% and 90% responses would occur in the later half of the assessment period $(T/2, T)$.

We compared the proposed multiple imputation methods (MI) to the standard method, complete method, and observed method which will be introduced here. For convenience, we refer to the latter three methods as SD, CP and OB, respectively. Basically, the simple Beta-Binomial model (4.1) is applied to estimate the posterior distribution of response rate for these three methods. For SD, we suspend the accrual and wait until the previously enrolled patients were fully followed-up. This method utilizes all the information and provides a benchmark for comparison. For the designs of CP and OB method, both recruit patients as the same rate with the proposed design using MI method. However, CP and OB methods only use partial information and lead to biased inferences resulting in terminating the trial inappropriately. Specifically, CP discards the missing responses and its inference is solely based on the outcomes of patients who have completed the assessment period or experienced responses during the assessment period. OB considers outcomes from all the treated patients, but it uses current outcomes to replace the final outcomes if they still have not been observed at the decision-making time. As mentioned in Section 2.2, CP and OB are both problematic with overestimation and underestimation of response rate respectively. For all four designs, continuous monitoring is conducted after n_0 patients have been fully followed-up in the trial. Therefore, we need to suspend the accrual after the n_0th patient enters the trial for designs of MI,

CP and OB. In our simulation studies, we set $n_0 = 5$.

The simulation results are shown in Table 4.1. We compared MI to SD, CP and OB in the terms of the average percentage of trial stopping, total number of patients assigned to the treatment and the total trial duration under each scenario.

Table 4.1: The percentage of trial stopping, total number of patients assigned to the treatment and the total trial duration under the standard, complete, observed and multiple imputation methods.

	percentage of stop				# of patients				trial duration			
$F_t(T)$	SD	CP	OB	MI	SD	CP	OB	MI	SD	CP	OB	MI
Response lower bound ϕ=0.3, 90% response in $(T/2, T)$, $A/I = 6$												
0.1	99.3	98.8	99.7	98.8	9.5	11.8	8.2	11.1	55.7	16.8	13.2	16.2
0.2	75.3	70.6	89.0	74.5	23.9	27.0	16.6	25.4	136.9	33.5	22.1	31.7
0.3	33.5	29.8	57.6	33.7	37.2	39.1	28.1	37.8	207.7	47.6	35.2	46.1
0.4	10.3	9.7	29.4	10.7	45.8	46.1	38.4	45.8	248.7	55.7	46.9	55.3
0.5	3.0	3.0	11.7	3.3	48.7	48.7	45.2	48.6	256.3	58.5	54.6	58.4
Response lower bound ϕ=0.4, 90% response in $(T/2, T)$, $A/I = 6$												
0.2	98.7	96.4	99.6	97.5	12.4	16.6	8.8	15.1	70.7	21.8	13.8	20.2
0.3	71.5	64.3	89.1	70.3	26.0	30.2	15.8	27.6	144.9	37.0	21.4	34.1
0.4	30.1	23.8	59.7	29.5	39.0	41.9	27.3	40.1	211.6	50.8	34.3	48.6
0.5	9.0	6.4	32.3	8.8	46.3	47.7	36.9	46.8	244.1	57.3	45.2	56.4
0.6	2.0	1.2	13.0	2.5	49.1	49.5	44.7	49.0	251.4	59.4	54.0	58.9
Response lower bound ϕ=0.5, 90% response in $(T/2, T)$, $A/I = 6$												
0.3	95.3	92.5	99.1	95.0	15.1	20.3	9.2	16.6	84.0	25.6	14.3	21.9
0.4	63.7	56.3	88.8	67.9	28.7	34.0	15.8	27.8	155.2	41.1	21.4	34.5
0.5	25.2	20.0	63.5	32.7	41.2	44.0	25.1	38.8	217.1	53.0	32.0	47.2
0.6	6.4	4.5	37.3	11.5	47.4	48.4	35.1	45.5	242.3	58.1	43.3	55.0
0.8	0.3	0.2	4.3	0.7	49.9	49.9	48.2	49.7	237.4	59.9	58.0	59.7
Response lower bound ϕ=0.3, 70% response in $(T/2, T)$, $A/I = 6$												
0.1	99.3	98.5	99.8	98.9	9.8	12.4	8.5	11.4	56.8	17.5	13.5	16.5
0.2	76.2	68.1	87.4	75.9	23.2	27.4	17.1	25.0	128.9	33.9	22.7	31.2
0.3	33.6	27.9	51.4	33.3	37.3	39.5	30.4	38.0	199.1	48.1	37.8	46.3
0.4	9.3	8.7	22.0	9.7	46.1	46.5	41.3	46.1	235.8	56.1	50.2	55.7
0.5	4.5	3.9	10.5	4.5	48.0	48.2	45.6	48.0	235.2	58.1	55.1	57.8
Response lower bound ϕ=0.4, 70% response in $(T/2, T)$, $A/I = 6$												
0.2	97.7	95.9	99.4	97.4	13.1	18.1	9.5	15.5	72.8	23.3	14.5	20.7
0.3	71.2	60.8	87.6	70.6	25.9	31.8	16.8	28.1	138.5	38.8	22.4	34.6
0.4	28.0	20.3	54.4	28.2	39.4	43.1	28.9	40.4	201.2	52.1	36.1	49.0
0.5	8.5	4.9	26.5	7.1	46.5	48.2	39.3	47.4	228.1	57.9	48.0	57.0
0.6	2.3	1.0	11.2	1.8	49.1	49.6	45.4	49.3	229.6	59.5	54.9	59.2
Response lower bound ϕ=0.5, 70% response in $(T/2, T)$, $A/I = 6$												
0.3	95.9	92.3	99.2	96.5	14.4	20.9	9.7	15.9	76.9	26.2	14.7	21.1
0.4	65.3	54.7	86.8	67.1	28.5	35.4	17.0	28.9	145.9	42.7	22.7	35.5
0.5	27.5	17.2	59.6	31.7	40.3	44.7	27.1	39.3	197.5	53.9	34.1	47.7
0.6	6.2	3.7	28.9	9.3	47.5	48.6	38.5	46.5	221.7	58.5	47.0	56.0
0.8	0.6	0.2	3.4	0.8	49.8	49.9	48.6	49.7	208.4	59.9	58.4	59.6

Table 4.1 continued.

	percentage of stop				# of patients				trial duration			
$F_t(T)$	SD	CP	OB	MI	SD	CP	OB	MI	SD	CP	OB	MI
Response lower bound ϕ=0.3, 50% response in $(T/2, T)$, $A/I = 6, K = 6$												
0.1	99.7	98.7	99.9	99.4	9.7	12.4	8.7	11.2	55.3	17.5	13.7	16.2
0.2	74.6	65.2	82.8	73.6	23.6	28.3	19.8	25.4	127.0	35.0	25.7	31.7
0.3	32.5	25.3	46.1	30.6	37.6	40.4	32.8	38.8	191.4	49.1	40.5	47.3
0.4	11.1	8.8	19.4	10.5	45.3	46.1	42.1	45.6	217.6	55.7	51.1	55.0
0.5	3.4	2.8	5.7	3.1	48.5	48.7	47.6	48.6	218.9	58.6	57.3	58.5
Response lower bound ϕ=0.4, 50% response in $(T/2, T)$, $A/I = 6, K = 6$												
0.2	97.2	93.8	98.8	96.3	12.9	18.6	10.2	15.7	69.8	24.0	15.3	20.9
0.3	68.6	55.7	81.4	66.5	26.5	33.8	19.8	29.5	135.0	41.0	25.7	36.2
0.4	28.2	19.5	47.7	26.5	39.4	43.2	31.4	40.8	189.0	52.2	39.1	49.5
0.5	8.1	4.6	21.6	7.9	46.7	48.1	41.3	46.9	211.4	57.9	50.2	56.5
0.6	3.3	1.9	9.6	2.3	48.6	49.2	45.9	49.0	205.4	59.1	55.5	58.9
Response lower bound ϕ=0.5, 50% response in $(T/2, T)$, $A/I = 6, K = 6$												
0.3	95.9	90.4	98.2	95.4	14.7	22.8	10.7	16.8	75.3	28.2	15.8	22.0
0.4	67.3	51.0	82.4	66.9	28.0	36.6	19.8	28.7	134.8	44.0	25.7	35.3
0.5	24.0	13.7	51.1	26.1	41.3	45.7	30.9	40.8	186.0	55.0	38.3	49.5
0.6	6.5	2.5	21.2	6.9	47.3	49.0	41.5	47.2	199.8	58.9	50.4	56.9
0.8	0.0	0.0	1.2	0.1	50.0	50.0	49.5	50.0	181.9	60.0	59.5	60.0

First, we considered high skewed late-onset response data, i.e., 90% of responses would occur in the later half of the assessment period $(T/2, T)$. For the scenario of the lower bound of response rate $\phi = 0.3$, if the true response rate $F_t(T) = 0.1$, the experimental treatment has lower efficacy and we should terminate the trial early to avoid assigning more patients to it. Under this scenario, SD terminates 99.3% trials for futility and assigns average 9.5 patients to the treatment. MI, CP and OB perform very similarly with SD. They terminate the trial for futility at the percentages of 98.8%, 98.8% and 99.7% and assign 11.1, 11.8 and 8.2 patients to the treatment, respectively. For SD, because of frequently accrual suspension, the trial duration is much longer than the other three methods. Considering the scenario with the same lower bound $\phi = 0.3$, if the true response rate $F_t(T) = 0.2$, we also should terminate the trial early and assign fewer patients to the treatment. Under this scenario, SD terminates 75.3% trials and assign 23.9 patients to the treatment. Our proposed MI method performs much more closer results with SD, which terminates 74.5% trials and assign 25.4 patients to the treatment. For CP, patients who would experience responses during the assessment period are more likely to be included for inference and therefore it overestimates the response rate which results in low percentage of early stopping. Hence, CP terminates the trials with the percentage of 70.6%, which is lower than that of SD, and assigns 27.0 patients to the treatment, which is higher than that of SD. OB considers the final responses of partially followed-up patients as no response. However, these patients might experience responses during the remaining assessment period. Therefore, it results in underestimation of the response rate and high percentage of early stopping. Comparing with SD, OB terminates the trials with lower percentage (89.0% versus 75.3%) and assigns fewer patients (16.6 versus 23.9) to the treatment. When the true response

rate $F_t(T) = 0.3$, SD terminates trials with the percentage of 33.5% and assigns 37.2 patients to the treatment. MI still performs better than CP and OB. Comparing with SD, its percentage of early stopping (33.5% versus 33.7%) and the number of patients assigned to the treatment (37.2 versus 37.8) are both very close to the results of SD. For the scenarios of considering true response rates $F_t(T) = 0.4$ and $F_t(T) = 0.5$, MI still has the best performance among all three methods, and has very comparable results with SD. However, the trial duration of SD is much longer than MI and results in a fatal implementation problem in the real trial. For all these five scenarios, CP outperforms than OB on average, since due to the high skewed late-onset responses, very few patients would experience responses at the early part of the assessment period which reduces the degree of overestimation for CP method. Similarly, we considered five scenarios with the lower bound $\phi = 0.4$ and another five scenarios with the lower bound $\phi = 0.5$. The same conclusion of MI outperforming CP and OB is made for all scenarios.

To further evaluate the performance of our proposed method, we considered different degrees of responses occurring in latter half of the assessment period, $(T/2, T)$. Similarly, we considered scenarios combining 70% of responses occurring in $(T/2, T)$ with different lower bounds. Taking the case of the lower bound $\phi = 0.4$ as an example, we listed 5 scenarios with different true response rates $F_t(T)$. On average, MI has very comparable results with SD and outperforms CP and OB. Specifically, considering $F_t(T) = 0.2$, MI terminates the trial for futility with the percentage of 97.4%, which is very close to the stopping percentage of SD (97.7%). When the true response rate is set at $F_t(T) = 0.6$ higher than the lower bound $\phi = 0.4$, SD terminates only 2.3% of trials for futility. Under this scenario, MI outperforms CP and OB again with stopping percentage at 1.8%. Meanwhile, MI

assigns similar number of patients to the experimental drug comparing with SD, but its trial duration is much shorter than the duration of SD. For the case of 50% of responses would occur in the later half of the assessment period $(T/2, T)$, MI still has very comparable performance with SD under many scenarios with different lower bounds and true response rates.

4.2.2 Sensitivity Analysis

To further evaluate the performance of MI, we generated the late-onset data from log-logistic distribution and compared the simulation results with SD, CP and OB. We considered two cases: in one case setting the lower bound ϕ at 0.4 with 90% responses occurring in $(T/2, T)$; in the other setting the lower bound ϕ at 0.5 with 70% responses occurring in $(T/2, T)$. The simulation results under these settings are displayed in Table 4.2. From the table, we made the same conclusion comparing with previous simulations which generate the time to response data from the Weibull distribution. Specifically, MI performs the comparable results with SD for both two cases and outperforms CP and OB. Taking the first case as an example, if the true response rate $F_t(T) = 0.3$, MI terminates the trial with 68.2% and assigns 28.1 patients to the treatment, which are very close to the results from SD (71.5% and 26.0).

We conducted sensitivity analyses to check the impact of parameter K and A/I ratio. The simulation results are displayed in Table 4.3. We considered the case of setting lower bound $\phi = 0.4$ and 90% responses occurring in $(T/2, T)$. The first five rows represent the analysis of checking the robustness of parameter K under different values of $F_t(T)$. Here, the A/I ratio is still set at 6, which is the same as the settings of previous simulation results. Specifically, we considered the scenarios with $K = 10, 12$ and compared the results with the results in Table 4.1, which set

$K = 6$ with the same ϕ and late-onset degree. From the results, we found similar simulation results for different values of K. For example, considering $F_t(T) = 0.2$, MI terminates the trial for futility with the percentages of $97.5\%, 96.8\%$ and 96.2% for $K = 6, 10, 12$, respectively. MI also assigns the similar numbers of patients to the treatment with the similar trial duration for different values of K. The results show that the number of intervals for piece-wise model has negligible effect on the results when it is set within a reasonable range.

Table 4.2: Sensitivity analysis for the proposed multiple imputation method with data generated from log-logistic distribution.

	percentage of stop				# of patients				trial duration			
$F_t(T)$	SD	CP	OB	MI	SD	CP	OB	MI	SD	CP	OB	MI
Time to event data is generated from log-logistic distribution												
Response lower bound ϕ=0.4, 90% response in $(T/2, T)$, $A/I = 6, K = 6$												
0.2	97.5	95.8	99.8	96.8	12.6	16.9	9.1	15.4	71.8	22.1	14.1	20.6
0.3	69.7	61.9	88.9	68.2	26.6	31.0	15.9	28.1	147.9	37.9	21.4	34.7
0.4	28.4	22.7	60.3	29.2	39.9	42.6	27.2	40.2	215.8	51.4	34.2	48.7
0.5	9.9	7.2	31.1	9.8	46.0	47.3	37.7	46.4	240.7	57.0	46.1	55.9
0.6	1.7	1.0	15.0	1.9	49.3	49.6	43.8	49.3	249.7	59.6	53.1	59.2
Response lower bound ϕ=0.5, 70% response in $(T/2, T)$, $A/I = 6, K = 6$												
0.3	95.6	91.9	99.1	95.8	15.3	21.9	10.1	16.7	81.8	27.3	15.2	21.9
0.4	63.3	53.8	84.5	65.3	29.1	35.2	18.0	29.2	148.5	42.5	23.8	36.0
0.5	25.8	17.1	57.9	29.4	40.7	44.7	27.7	39.8	199.1	53.8	34.8	48.4
0.6	4.9	2.2	27.8	7.0	48.0	49.2	39.1	47.2	223.0	59.1	47.7	56.9
0.8	0.2	0.1	2.7	0.2	49.9	50.0	48.9	49.9	207.0	60.0	58.8	59.9

The last five rows in Table 4.3 represent the analysis of checking the impact of A/I ratio. For comparison, the parameter K is set at 6, which is the same with the simulations in Table 4.1. We considered $\phi = 0.4$ and 90% late-onset degrees. We considered the A/I ratio at larger values of 8, 12 and compared the results in Table 4.1, which set A/I ratio at 6 with the same ϕ and late-onset degree. The larger value of A/I ratio indicates faster accrual rate and higher percentage of missing responses at the decision-making time. Therefore, it leads to a difficult case when

A/I ratio increases. From the simulation results, we found that with higher A/I ratio, MI incorrectly terminates the trial with a higher percentage for most scenarios. Considering a scenario with true response rate $F_t(T) = 0.6$, the stopping percentages for A/I ratio at 6, 8, 12 are 2.5, 3.6, 3.9, respectively.

Table 4.3: Sensitivity analysis for the proposed multiple imputation method with different values of K and different values of A/I ratio

$F_t(T)$	% of stop	# of patients	trial duration	% of stop	# of patients	trial duration
		K=10			K=12	
0.2	96.8	15.3	20.5	96.2	15.6	20.8
0.3	69.1	28.0	34.6	69.0	27.9	34.5
0.4	29.9	40.0	48.5	31.7	39.2	47.6
0.5	9.1	46.5	56.1	8.7	46.7	56.3
0.6	1.9	49.2	59.1	2.6	49.0	58.8
		A/I=8			A/I=12	
0.2	97.7	15.6	17.1	97.1	16.8	14.0
0.3	69.2	28.3	28.1	70.0	28.8	21.5
0.4	30.3	40.0	38.9	32.4	40.0	29.2
0.5	11.4	45.7	44.2	9.9	46.6	33.7
0.6	3.6	48.6	46.7	3.9	48.5	35.0

CHAPTER 5

Conclusions

To account for the unique properties of biological agents, we proposed a new Bayesian phase I/II design for trials that evaluate combinational biological agents. A change-point model is used to capture the feature that the dose-toxicity surface of biological agents may plateau at high dose levels, and a second-order logistic model is employed to accommodate non-monotonic patterns for the dose-efficacy relationship. We proposed a novel dose-finding algorithm that adaptively encourages the exploration of two-dimensional dose-toxicity and dose-efficacy surfaces during dose finding. In the early stage of the trial, the algorithm gives higher priority to trying new doses, and toward the end of the trial it assigns patients to the most effective dose that is safe. Extensive simulations show that the proposed design has good operating characteristics with a high probability of selecting the BODC. The advantage of our proposed design over the greedy design further verifies the importance of the dose-exploration algorithm incorporated in our design.

The proposed design is appropriate for trials in which toxicity and efficacy outcomes are observed quickly. If toxicity and particularly efficacy cannot be ascertained in a timely manner, the proposed design may be less useful. To handle delayed toxicity and efficacy outcomes, we can extend our approach by modeling toxicity and

efficacy as time-to-event outcomes to accommodate censored observations [22]. In addition, in the proposed design, we are interested in finding the dose with highest efficacy and tolerable toxicity as the target BODC. Our design can be easily extended to the case that the target BODC is defined by a certain toxicity-efficacy trade-off function. In that case, the main exercise is to elicit a reasonable toxicity-efficacy trade-off (or utility) function from clinicians [47]. Once the trade-off is defined, our design can be directly applied by replacing efficacy with the trade-off as the criteria of dose escalation and selection.

To more efficiently handle the large number of combination therapies that must be tested, we proposed a Bayesian adaptive phase II screening design for trials combining multiple agents. Rather than testing each of the combinations independently, our design encompasses all the combinations of interest in a large screening trial. We model the main and synergistic effects of the treatment agents using a linear model and cast the screening problem into a Bayesian hypothesis testing problem. By using a factorial lead-in phase, we are able to quickly obtain preliminary estimates that each treatment combination is optimal, which enables us to quickly move into the adaptive phase of the algorithm. We utilize the encompassing prior with non-local constraints to accommodate the complex parameter constraints imposed by the hypotheses, and we continuously update the posterior probability that each treatment is best. Based on this posterior probability, we adaptively allocate patients to effective combinations and select the best treatment. The proposed design substantially outperformed a complete factorial design. Our design allocates more patients to better treatments while at the same time providing higher power to identify the best treatment at the end of the trial.

To address the unobserved responses for late-onset responses at the decision

making time, we proposed a Bayesian single-arm phase II design for estimating the response rate of the experimental drug. We conduct continuous monitoring to terminate the trial early for futility and avoid assigning unacceptable number of patients to inefficacious treatments. We handle the missing responses using the multiple imputation approach by modeling the hazard function of time to response data using a piece-wise exponential model. Extensive simulations show that the proposed design yields a desirable operating characteristics for different physician-specified lower bounds of response rate with different true response rates. The proposed design dramatically reduces the total length of the trial duration.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Elja Arjas and Dario Gasbarra. Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, 4:505–524, 1994.

[2] Helen Aslanidou, Dipak K. Dey, and Debajyoti Sinha. Bayesian analysis of multivariate survival data using Monte Carlo methods. *Canadian Journal of Statistics*, 26:38–48, 1998.

[3] James Babb, Andre Rogatko, and Shelemyahu Zacks. Cancer phase I clinical trials: Efficient dose escalation with overdose control. *Statistics in Medicine*, 17:1103–1120, 1998.

[4] B Nebiyou Bekele, Yuan Ji, Shen Yu, and Peter F Thall. Monitoring late-onset toxicities in phase I trials using predicted risks. *Bistatistics*, 9:442–457, 2008.

[5] Thomas M Braun, Peter F Thall, Hoang Nguyen, and Marcos de Lima. Simultaneously optimizing dose and schedule of a new cytotoxic agent. *Clinical Trials*, 4:113–124, 2007.

[6] Thomas M Braun and Shufang Wang. A hierarchical Bayesian design for phase I trials of novel combinations of cancer therapeutic agents. *Biometrics*, 66:805–812, 2010.

[7] George Casella and Edward George. Explaining the Gibbs sampler. *The American Statistician*, 46:167–174, 1992.

[8] Ying Kuen Cheung and Rich Chappell. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*, 56:1177–1182, 2000.

[9] Ying Kuen Cheung and Peter F. Thall. Monitoring the rates of composite events with censored data in phase II clinical trials. *Biometrics*, 58:89–97, 2002.

[10] Mark R Conaway, Stephanie Dunbar, and Shyamal D Peddada. Designs for single- or multiple-agent phase i trials. *Biometrics*, 60:661–669, 2004.

[11] Meichun Ding, Gary L. Rosner, and Peter Mller. Bayesian optimal design for phase II screening trials. *Biometrics*, 64:886–894, 2008.

[12] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis. Second Edition.* Chapman and Hall, 2003.

[13] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2:1360–1383, 2008.

[14] W.R. Gilks, N.G. Best, and K.K.C. Tan. Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 44:455–472, 1995.

[15] Steven N. Goodman. Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, 130:1005–1013, 1999.

[16] Ted A Gooley, Paul J Martin, Lloyd D Fisher, and Mary Pettinger. Simulation as a design tool for phase I/II clinical trials: an example form bone marrow transplantation. *Controlled Clinical Trials*, 15:450–462, 1994.

[17] Paulo M Hoff and Lee M Ellis. Targeted therapy trials: approval strategies, target validation, or helping patients? *Journal of Clinical Oncology*, 25:1639–1641, 2007.

[18] Nathan T Ihle and Garth Powis. Take your PIK: phosphatidylinositol 3-kinase

inhibitors race through the clinic and toward cancer therapy. *Molecular Cancer Therapeutics*, 8:1–9, 2009.

[19] Valen E. Johnson and John D. Cook. Bayesian design of single-arm phase II clinical trials with continuous monitoring. *Clinical Trials*, 6:217–226, 2009.

[20] Valen E. Johnson and Tito R. Mendoza. *Cancer Symptom Science: Measurement, Mechanisms, and Management*, chapter Bayesian adaptive design: a novel approach to test the effectiveness of symptom-reducing agents using patient-reported outcomes, pages 293–303. Cambridge University Press, 2010.

[21] Valen E. Johnson and David Rossell. On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72:143–170, 2010.

[22] John D. Kalbfleisch and Ross L. Prentice. *The statistical analysis of failure time data, 2nd edition.* New York: Wiley, 2002.

[23] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.

[24] Irene Klugkist and Herbert Hoijtink. The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, 51:6367–6379, 2007.

[25] Irene Klugkist, Bernet Kato, and Herbert Hoijtink. Bayesian model selection using encompassing priors. *Statistica Neerlandica*, 59:57–69, 2005.

[26] Edward L. Korn and Richard Simon. Using the tolerable-dose diagram in the design of phase I combination chemotherapy trials. *Journal of Clinical Oncology*, 11:794–801, 1993.

[27] Andrew Kramar, Anne Lebecq, and Emmanuel Candalh. Continual reassess-

ment methods in phase I trials of the combination of two drugs in oncology. *Statistics in Medicine*, 18:1849–1864, 1999.

[28] J Jack Lee, Xuemin Gu, and Suyu Liu. Bayesian adaptive randomization designs for targeted agent development. *Clinical Trials*, 0:1–13, 2010.

[29] J Jack Lee and Diane D Liu. A predictive probability design for phase II cancer clinical trials. *Cli*, 5:93–106, 2008.

[30] Roderick J.A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data (2nd ed.)*. New York: Wiley, 2002.

[31] Sumithra J Mandrekar, Yue Cui, and Daniel J Sargent. An adaptive phase I design for identifying a biologically optimal dose for dual agent drug combinations. *Statistics in Medicine*, 26:2317–2330, 2007.

[32] Sumithra J Mandrekar and Daniel J Sargent. Clinical trial designs for predictive biomarker validation: Theoretical considerations and practical challenges. *Journal of Clinical Oncology*, 27:4027–4034, 2009.

[33] John O'Quigley, Michael D Hughes, and Terry Fenton. Dose-finding designs for HIV studies. *Biometrics*, 57:1018–1029, 2001.

[34] John O'Quigley and Xavier Paoletti. Continual reassessment method for ordered groups. *Biometrics*, 59:430–440, 2003.

[35] John O'Quigley, Margaret Pepe, and Lloyd Fisher. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics*, 46:33–48, 1990.

[36] John O'Quigley and S Zohar. Experimental designs for phase I and phase I/II dose-finding studies. *British Journal of Cancer*, 94:609–613, 2006.

[37] Zuqiang Qiou, Nalini Ravishanker, and Dipak K. Dey. Multivariate survival analysis with positive stable frailties. *Biometrics*, 55:637–644, 1999.

[38] Mark J Ratain and Daniel J Sargent. Optimising the design of phase II oncology trials: The importance of randomisation. *European Journal of Cancer*, 45:275–280, 2009.

[39] David Rossell, Peter Mller, and Gary L. Rosner. Screening designs for drug development. *Biostatistics*, 8:595–608, 2007.

[40] Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. New, 1987.

[41] Daniel J. Schaid, Sam Wieand, and Terry M. Therneau. Optimal two-stage screening designs for survival comparisons. *Biometrika*, 77:507–513, 1990.

[42] Richard Simon and Edward L. Korn. Selecting drug combinations based on total equivalent dose (dose intensity). *Journal of the National Cancer Institute*, 82:1469–1476, 1990.

[43] Richard M. Simon. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*, 10:1–10, 1989.

[44] Sonali M Smith, Koen Besien, Theodore Karrison, Janet Dancey, Peter Mclaughlin, Anas Younes, Scott Smith, Patrick Stiff, Eric Lester, Sanjiv Modi, L Austin Doyle, Everett E Vokes, and Barbara Pro. Temsirolimus has activity in nonmantle cell non-Hodgkins lymphoma subtypes: The University of Chicago phase II consortium. *Journal of Clinical Oncology*, 28:4740–4746, 2010.

[45] Quentin F. Stout and Janis Hardwick. Optimal screening designs with flexible cost and constraint structures. *Journal of Sequential Planning and Inference*, 132:149–162, 2005.

[46] Mario Stylianou and Nancy Flournoy. Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics*, 58:171–177, 2002.

[47] Peter F Thall and John D Cook. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*, 60:684–693, 2004.

[48] Peter F Thall, RandAll E Millikan, Peter Mueller, and Sang-Joon Lee. Dose-finding with two agents in phase I oncology trials. *Biometrics*, 59:487496, 2003.

[49] Peter F. Thall and Kathy E. Russell. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics*, 54:251–264, 1998.

[50] Peter F. Thall, Richard Simon, and Susan S. Ellenberg. Two-stage selection and testing designs for comparative clinical trials. *Biometrika*, 75:303–310, 1988.

[51] Peter F. Thall, Richard M. Simon, and Elihu H. Estey. New statistical strategy for monitoring safety and efficacy in single-arm clinical trials. *Journal of Clinical Oncology*, 14:296–303, 1996.

[52] Nolan A Wages, Mark R Conaway, and John O'Quigley. Continual reassessment method for partial ordering. *Biometrics*, 67:1555–1563, 2011.

[53] Kai Wang and Anastasia Ivanova. Two-dimensional dose finding in discrete dose space. *Biometrics*, 61:217–222, 2005.

[54] You-Gan Wang and Denis Heng-Yan Leung. An optimal design for screening trials. *Biometrics*, 54:243–250, 1998.

[55] J. Kyle Wathen, Peter F. Thall, John D. Cook, and Elihu H. Estey. Accounting for patient heterogeneity in phase II clinical trials. *Statistics in Medicine*, 27:2802–2815, 2008.

[56] Tzy-Jyun Yao, Colin B. Begg, and Philip O. Livingston. Optimal sample size for a series of pilot trials of new agents. *Biometrics*, 52:992–1001, 1996.

[57] Tzy-Jyun Yao and E. S. Venkatraman. Optimal two-stage design for a series of pilot trials of new agents. *Biometrics*, 54:1183–1189, 1998.

[58] Guosheng Yin, Yisheng Li, and Yuan Ji. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics*, 62:777–787, 2006.

[59] Guosheng Yin and Ying Yuan. Bayesian model averaging continual reassessment method in phase I clinical trials. *Journal of the American Statistical Association*, 104:954–968, 2009.

[60] Ying Yuan and Guosheng Yin. Sequential continual reassessment method for two-dimensional dose finding. *Statistics in Medicine*, 27:5664–5678, 2008.

[61] Ying Yuan and Guosheng Yin. Bayesian phase I/II adaptively randomized oncology trials with combined drugs. *The Annals of Applied Statistics*, 5:924–942, 2011.

[62] Xian Zhou, Suyu Liu, Edward S Kim, Roy S. Herbst, and J Jack Lee. Bayesian adaptive design for targeted therapy development in lung cancer - a step toward personalized medicine. *Clinical Trials*, 5:181–193, 2008.

VITA

Chunyan Cai was born in Feidong, Anhui, China on October 15th, 1986, the Daughter of Chuanfa Cai and Jiayun Zhang. She graduated from Feidong No.1 High School in 2003. She attended the University of Science and Technology of China in Hefei, Anhui, China and received her Bachelor of Science degree in Biology in 2007. In August of 2007, she entered the University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences. She hopes to receive her Doctor of Philosophy degree in Biostatistics in August 2012.