Generalized Statistical Approaches for the Design for Phase I Trials

by

Nan Jia

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2012**

Doctoral Committee:

Associate Professor Thomas M. Braun, Chair
Associate Professor Timothy D. Johnson
Assistant Professor Lu Wang
Professor Edward D. Rothman

For my parents, my husband, and my son

ACKNOWLEDGEMENTS

First and foremost, I want to express my deepest gratitude to my thesis and academic advisor, Dr. Thomas Braun. All through my research years here at the department of Biostatistics, Dr. Braun has always been very supportive with his infinite patience and broad knowledge. It was Dr. Braun who encouraged me to explore the field of clinical trials, generously shared his valuable advice, familiarized me to the academic atmosphere, and also trained me on scholarly article writing. I greatly appreciate the efforts that Dr. Braun had made toward my gaining the knowledge and techniques that I did not originally have. Without Dr. Braun′s constant guidance and encouragement, this work will not be what it is.

I would like to thank Dr. Edward Rothman, who employed me as a Graduate Student Statistical Consultant at the Center for Statistical Consultation and Research (CSCAR). During my two-and-a-half year appointment with CSCAR, I have learned many new statistical techniques, become more experienced and knowledgeable with many statistical procedures and software, and most importantly, acquired much experience in handling various projects from a wide range of backgrounds. I also thank Dr. Rothman for serving on my dissertation committee as the cognate member.

I thank Dr. Timothy Johnson for sharing his opinions on my dissertation from a Bayesian point of view. Dr. Johnson has been very helpful not only in making good suggestions on my thesis work, but also in offering excellent classes that I enjoyed very

much. I also want to thank Dr. Lu Wang for serving on my dissertation committee and providing valuable advice on my thesis.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vii

LIST OF TABLES viii

ABSTRACT xi

CHAPTER

I. Introduction 1

1.1 Background 1
1.2 Significance of Research 2

II. The Accelerated Adaptive Biased Coin Design 6

2.1 Introduction 6
2.2 Methods 10
2.2.1 The traditional BCD and the accelerated BCD . . . 10
2.2.2 The Adaptive Accelerated Biased Coin Design . . . 13
2.3 Simulation Studies 16
2.3.1 Description of Simulation Settings 16
2.3.2 Simulation Results 18
2.4 Conclusion and Discussion 21

III. A Generalized Continual Reassessment Method for Two-Agent Phase I Trials 28

3.1 Introduction 28
3.2 Methods 31
3.2.1 Model Description 31

3.2.2 Deriving Hyperparameter Values and Rescaled Doses 33
3.2.3 Trial Conduct 35
3.3 Simulation Studies 37
3.3.1 Simulation results of the gCRM 37
3.3.2 Comparison to Existing Approaches 41
3.3.3 Assessing Sensitivity to Elicited DLT Probabilities . 44
3.4 Discussion 45

IV. A Time-To-Event Approach to the gCRM 58

4.1 Introduction 58
4.2 Methods 61
4.2.1 Description of the gCRM 61
4.2.2 Incorporation of Partial Follow-up 62
4.2.3 Trial Conduct 68
4.3 Simulation Studies 69
4.3.1 General Description 69
4.3.2 When Weights are Based on Actual Distribution of DLT Times 70
4.3.3 When Weights are Based on Assumed Distribution of DLT Times 73
4.4 Concluding Remarks 75

V. Summary and Future Work 84

BIBLIOGRAPHY 87

LIST OF FIGURES

Figure

3.1 Schematic representation of Phase I trial of two agents, with four doses of each agent. 46

3.2 Average dose assigned to each patient in simulations. 47

LIST OF TABLES

Table

2.1 True DLT probabilities in scenarios examined in simulations. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.) 23

2.2 Proportion of 2,000 simulations in which each dose was identified as the MTD when the doses have unequally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose that was selected most often on average. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.) 24

2.3 Average percentage of patients assigned to each dose in 2,000 simulations when the doses have unequally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose most often assigned to patients. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.) 25

2.4 Proportion of 2,000 simulations in which each dose was identified as the MTD when the doses have equally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose that was selected most often. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.) 26

2.5 Average percentage of patients assigned to each dose in 2,000 simulations when the doses have equally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose most often assigned to patients. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.) 27

3.1 Operating characteristics of the gCRM in seven hypothetical scenarios. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate that is within 10 points of the desired DLT rate $\Gamma = 0.20$. 48

3.2 Comparison of gCRM operating characteristics to those in Braun and Wang (2010) in three scenarios. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate within 10 points of the desired DLT rate $\Gamma = 0.20$. . 49

3.3 Comparison of the operating characteristics of the gCRM to the designs of Yin and Yuan (2009a; 2009b) and the POCRM of Wages, Conaway, O'Quigley (2011a). All values are percentages multiplied by 100. Boldfaced numbers correspond to dose combinations with a DLT rate equal to the desired DLT rate. 50

3.4 Operating characteristics of the gCRM using three different sets of elicited DLT probabilities. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate within 10 points of the desired DLT rate $\Gamma = 0.20$. 51

3.5 Operating characteristics of the gCRM when assuming the priors of the hyperparameters are uniformly distributed in their respective recommended range. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate that is within 10 points of the desired DLT rate $\Gamma = 0.20$. Simulation size is 1000, draw size is 5000 with a burn-in of 1000. 52

3.6 Simulation results of the gCRM using the power model. True DLT probabilities, and simulation results for scenarios A through F. All values are multiplied by 100. Bold faced numbers correspond to dose combinations with a true DLT rate exactly equal to the target DLT rate. 53

3.7 True DLT probabilities, and simulation results for scenarios with four levels of agent A and three levels of agent B. All values are multiplied by 100. Scenario A1 boundary DLT rates are used in all seven scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$. 54

3.8 True DLT probabilities, and simulation results for scenarios with five levels of agent A and three levels of agent B. All values are multiplied by 100. Scenario A2 boundary DLT rates are used in all seven scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$. 55

3.9 True DLT probabilities, and simulation results for scenarios with five levels of agent A and four levels of agent B. All values are multiplied by 100. Scenario A3 boundary DLT rates are used in all seven scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$. A sample size of 35 patients is used in all seven scenarios. 56

3.10 Simulations results of different combinations of μ, θ and σ^2 using the proportional odds logistic model of the gCRM. All values are multiplied by 100. Scenario A DLT rates are assumed as the true DLT rates in all the 11 scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$. Simulation size is 2000, draw size is 5000 with a burn-in of 2000. A sample size of 35 patients is used in all seven scenarios. 57

4.1 Operating characteristics of the wgCRM in Scenario A using 10 different weights. All values are multiplied by 100. 77

4.2 Operating characteristics of the wgCRM in Scenario B using 10 different weights. All values are multiplied by 100. 78

4.3 Operating characteristics of the wgCRM in Scenario C using 10 different weights. All values are multiplied by 100. 79

4.4 Operating characteristics of the wgCRM in Scenario D using 10 different weights. All values are multiplied by 100. 80

4.5 Operating characteristics of the wgCRM in Scenario E using 10 different weights. All values are multiplied by 100. 81

4.6 Operating characteristics of the wgCRM in Scenario F using 10 different weights. All values are multiplied by 100. 82

4.7 Operating characteristics of the wgCRM in Scenario G using 10 different weights. All values are multiplied by 100. 83

ABSTRACT

Generalized Statistical Approaches for the Design for Phase I Trials

by

Nan Jia

Chair: Thomas M. Braun

This research examines new design approaches for Phase I clinical trials, which are designed to study doses of the same agent or dose combinations of two agents in a small group of patients to determine the maximum tolerated dose or dose combination.

Our first focus is to propose an adaptive accelerated Biased Coin Design (aaBCD) that generalizes the traditional BCD design algorithm by incorporating an adaptive weight function based upon the amount of follow-up of each enrolled patient, so that the dose assignment for each eligible patient can be determined immediately with no delay, leading to a shorter trial overall.

We later focus on a generalized version of the Continual Reassessment Method (CRM), denoted gCRM, for identifying the maximum tolerated dose combination of two agents. For each dose of one agent, we apply the traditional CRM to study doses of the other agent; each of these CRM designs assumes the same dose-toxicity model, as well as the value of the parameter used in the model. However, each model includes a second parameter that varies among the models in an effort to

allow flexibility when modeling the probability of dose-limiting toxicity (DLT) of all combinations, yet borrow strength among neighboring combinations as well.

We lastly extend the gCRM by incorporating results for patients with incomplete follow-up into the decision rule for the assignment of a dose combination to the next available patient. We derive methods that account for the differing amounts of follow-up that could occur for the two agents and propose the use of a copula function to adjust for early- or late-onset DLTs. We show an optimal weight via theory and simulations that when the DLT times are distributed the same as assumed model, the optimal weight performs best among all the weight functions under consideration.

CHAPTER I

Introduction

1.1 Background

In drug development, clinical trials are a set of procedures that are conducted to assess drug safety, drug associated adverse events, and efficacy of new drugs as compared to standard therapy or placebo. Clinical trials are usually conducted in four Phases, where the aims respectively are: initial safety study to determine safe doses; initial analysis of efficacy to establish a final dose to pursue further; final testing to study the efficacy of a drug; and post-marketing surveillance. The number of patients recruited increases as the the trial advances to higher phases, and the entire study of an agent from Phase I to Phase IV can take years or even more than a decade.

Phase I trials are the first stage of testing in human subjects. One hallmark of most Phase I trials is that a relatively small number of patients will be studied, as investigators would prefer to expend patient resources in later Phase II and III trials examining the effectiveness of the agent. As a result, an appropriate Phase I trial design must be one that can identify the MTD accurately in a small number of patients as well as one that can treat as many patients at the MTD as possible.

Phase I trials normally include dose escalation studies, so that the best and safest dose can be found and to discover the point at which a compound has reached the highest allowable toxicity rate that is pre-specified by the investigators.

Distinguished by the number of agents involved, a Phase I trial can be categorized as either a one-agent trial or two-agent trial. A one-agent Phase I trial, also called the dose-escalation study, searches in a set of candidate doses of the same agent for a target dose whose DLT rate is closest to the pre-specified toxicity rate. A two-agent Phase I trial, which has two agents that each contains at least two doses, studies the dose-toxicity profile of the dose combinations, where every dose combination represents the combined treatment of one dose from each of the two agents. Therefore, a two-agent Phase I trial aims to determine which dose combination leads to a DLT rate closet to the target rate .

Every patient or cohort of patients is given a dose or dose combination and are observed for a period of time. The follow-up period after a patient has received their treatment usually lasts until several half-lives of the drug have passed, or until the patient has shown an adverse event, which ever comes first. The dose-toxicity profile is adaptively built based on the observed DLT outcomes of the currently enrolled patients, and the dose assignment for a future patient or cohort of patients, as well as the MTD, is determined accordingly.

1.2 Significance of Research

This dissertation aims to address two issues in currently existing Phase I trials. The first issue is, although not always explicitly stated, most existing Phase I designs require that the current patient or group of patients must complete their follow-up

before the next available patient can be assigned with a dose or a combination of doses. In a situation where the follow-up period is relatively long as compared to the average patient inter-arrival time, the length of the whole trial can be impractically long, and it is not ethical to let a available patient wait for a long period before he can be assigned with a treatment. Furthermore, for the two agent trials, where the goal is to find the maximum tolerated combination (MTC) among multiple doses of each agent, no methodology exists for incorporating partial follow-up of patients.

Another issue in Phase I trials is that, the two-agent Phase I trials, while being more and more popular within medical researchers, are not as well developed as the one-agent trials. Most of the existing two-agent Phase I trials suffer from various disadvantages that includes requiring too much information from the investigator, being sensitive to model mis-specifications, not allowing both agents to vary simultaneously when assigning a treatment to a future patient, etc.

In an effort to solve the first issue, in Chapters II and IV of this dissertation, for one agent and two agents trials respectively, we aim to utilize the partial information collected on the patients that are still under follow-up upon the arrival of next available patient along with the information from the fully followed patients, to assign a dose or combination of doses to the next available patient immediately, thereby significantly reduce the length of trial.

As an approach to the second issue, in Chapter III, we aim to develop a two-agent dose-finding algorithm that, without dramatically increasing the sample size from the one-agent trials, achieves high accuracy in targeting the MTC while maintaining patient safety, and performs well even when the information from the investigators largely deviates from the truth, and is robust to model mis-specifications.

In more details, Chapter II focuses on modifying the traditional Biased Coin

Design (*Durham and Flournoy*, 1995a) when partial follow-up exists. The traditional Biased Coin Design (BCD) is a non-parametric up-and-down design that aims to target the maximum tolerated dose (MTD) out of several doses of an agent. The BCD requires that the most recently enrolled patient must has been completely followed before the next available patient can be assigned with a dose. This may introduce some practical issues. For example, the next available patient is seriously diseased thus is not able/unwilling to wait; or the total length of the trial is impractically long as compared to the average patient inter-arrival times. We develop a method called the accelerated adaptive Biased Coin Design (aaBCD) for incorporating the partial contributions of the incomplete follow-ups into the trial, so that a patient can be entered into a trial immediately at their arrival regardless of whether the previously enrolled patients are completely followed. The patient waiting time is eliminated and the duration of the whole trial can be greatly shortened by the aaBCD. We demonstrate via simulations that even with prompt patient accrual, the aaBCD is able to maintain comparable accuracy as the BCD in terms of both the MTD selection and the patient assignments.

In Chapter III, we focus on a parametric dose-finding algorithm for two agents. We propose a generalized version of the Continual Reassessment Method (CRM), denoted gCRM, for identifying the MTC of two agents. For each dose of one agent, we apply the traditional CRM to study doses of the other agent; each of these CRM designs assumes the same dose-toxicity model, as well as the value of the parameter used in the model. However, each model includes a second parameter that varies among the models in an effort to allow flexibility when modeling the probability of dose-limiting toxicity of all combinations, yet borrow strength among neighboring combinations as well. We incorporate an adaptive Bayesian algorithm to sequentially

assign each patient to the most appropriate dose combination, as well as focus patient assignments to a dose combination that has a DLT probability closest to a pre-specified target rate. We test the performance of our method via extensive simulations in various scenarios that are likely to arise in two-agent Phase I trials. We also compare the operating characteristics of our approach to several existing methods that were recently published.

In Chapter IV we focus on the development of a time-to-event version of the gCRM. We develop weight functions that incorporate the partial information from subjects with uncomplete follow-up into the likelihood function when determining the most appropriate dose combination for the next available patient. We consider nine different weight functions and compare them to the gCRM when information from the partial follow-up is ignored. We show via simulation studies that the weighted gCRM can handle a great amount of partial follow-up without significantly reducing the overall performance of the algorithm. In addition, one of the nine weight functions is shown to perform the best, in terms of identifying the MTC and assigning patients to the MTC.

CHAPTER II

The Accelerated Adaptive Biased Coin Design

2.1 Introduction

One agent phase I clinical trials, also known as dose-finding or dose-escalation studies, are designed to determine which dose of an agent can be given to patients before an unacceptable fraction, Γ, of patients experience dose-limiting toxicities (DLTs). The largest dose studied that maintains a DLT fraction closest to Γ is defined as the maximum tolerated dose (MTD).

Existing one agent phase I trial designs can be categorized in a number of ways. *O'Quigley and Zohar* (2006) define trial designs as being either "memoryless" in which the dose assignment for the next patient or cohort of patients is determined solely from the DLT experience of the most-recently enrolled patient or cohort of patients, or "with memory", in which the dose assignment of each patient or cohort of patients is determined from a model incorporating the DLT experience of all the enrolled patients.

Most memoryless designs are also known as "up-and-down designs", one variant of which is the widely-used 3+3 method (*Storer*, 1989). In the 3+3 method, patients are enrolled in cohorts of three, with each member of a cohort assigned to the same

dose. If no patients in a cohort experience DLT, the next cohort is assigned to the next higher dose. If two or three patients in a cohort experience DLT, the trial is terminated, with the MTD defined at one dose below the assignment of the current cohort. If one member of the cohort experiences DLT, another cohort of three patients is assigned to the same dose. If no additional DLTs occur, the next cohort is assigned to the next higher dose. However, if at least one additional DLT occurs, the trial is terminated with the MTD defined at one dose below the assignment of the current cohort. Although the 3+3 design is widely used because it is easy to administer, numerous authors, such as *Ahn* (1998), *Storer* (2001), and *Korn et al.* (1994), have shown that the 3+3 design tends to both identify the MTD at, and assign a large percentage of patients to, doses below the MTD.

Due to the poor performance of the 3+3 method, statistical researchers began to propose the use of one agent phase I trial designs "with memory", the first of which was the Continual Reassessment Method (CRM) of *O'Quigley et al.* (1990), with later modifications proposed by *Faries* (1994) and *Goodman et al.* (1995). The CRM assumes the probability of DLT is related to dose d through a parametric function $F(\beta; d)$ that is monotonic in dose. Given a desired probability of DLT, Γ, the CRM attempts to identify which of the doses under study has $F(\beta; d)$ closest to Γ. A prior distribution is placed on β and the posterior mean of β is sequentially updated as each enrolled subject completes their follow-up. The first patient is usually assigned to the lowest dose, and patient $i, i = 2, \ldots, N$ is assigned to the dose in which $F(\tilde{\beta}_{i-1}; d)$ is closest to Γ, where $\tilde{\beta}_{i-1}$ is the posterior mean of β based upon the data from the first $(i-1)$ patients. The MTD is the dose, d, for which $F(\tilde{\beta}_N; d)$ is closest to Γ. A maximum-likelihood version of the CRM for estimation of β also exists (*Shen and O'Quigley*, 1996).

Although the 3+3 method suffers from poor operating characteristics, this is not a general property of all up-and-down phase I trial designs. The Biased Coin Design (BCD) of *Durham and Flournoy* (1995a) is a valid competitor to the CRM, although the BCD is rarely used in practice. In its basic form, the BCD assigns each new patient to a dose depending solely upon the DLT outcome of the current patient. Specifically, if the current patient has a DLT, the next subject is assigned randomly to either the same dose or the next lowest dose, with the probability of either assignment dependent upon Γ. If the current patient does not have a DLT, the next patient is assigned to the next highest dose. *Durham and Flournoy* (1995b), *Stylianou and Flournoy* (2002), and *Giovagnoli and Pintacuda* (1998) have all published work demonstrating the excellent operating characteristics of the BCD.

The 3+3 method, CRM, and BCD all share one limitation that fails to address the needs of practical phase I clinical trials: they all require the complete follow-up of all enrolled patients before a new patient (or cohort of patients) can be enrolled. The creators of the CRM suggested that, in the absence of complete follow-up of the current patient, all patients should be enrolled at whatever dose the current patient received. A similar approach was suggested for the BCD, called the accelerated Biased Coin Design (ABCD), by *Stylianou and Follmann* (2004). However, the idea of making dose assignments based on the last-completed patient is still inadequate in studies with rapid patient accrual, as many patients could still be treated at sub-optimal doses. The rolling six design (RSD) is a recent modification to the 3+3 design that has been shown to shorten the duration of Phase I trials relative to the 3+3 design (See *Skolnik et al.* (2008) and *Onar-Thomas and Xiong* (2010)). However, like the 3+3 method, *Skolnik et al.* (2008)demonstrates that the RSD continues to identify the MTD and assign too many subjects to sub-optimal doses.

The time-to-event CRM (TITE-CRM) (*Cheung and Chappell*, 2000) is the first phase I trial design that can determine a dose assignment for each eligible patient using all data collected in the trial, including the information for subjects with incomplete follow-up. The TITE-CRM uses the same adaptive, Bayesian methods of the original CRM. However, when computing the posterior mean for β, a weight $w(t, \tau)$ for each patient with incomplete follow-up is used in the likelihood, where $0 \leq t \leq \tau$ is the amount of time the patient has been followed, and τ is the maximum amount of follow-up planned for each patient. The weighted likelihood can actually be derived from a cure model (*Braun*, 2006), showing that the appropriate weight function should be the cumulative distribution (CDF) of DLT times for those experiencing DLT by τ. The standard form of the TITE-CRM uses the weight $w(t, \tau) = t/\tau$, which assumes uniform DLT times, although any plausible CDF could be used if non-uniformly distributed DLTs were expected. *Braun* (2006) and *Bekele et al.* (2008) have proposed methods to make the form of the weights used by the TITE-CRM adaptive to the actual DLT times observed in the trial.

This work focuses upon an adaptive accelerated BCD (aaBCD), a design in which the BCD is generalized so that patients can be enrolled as soon as they are eligible, thereby preventing delayed treatment of patients and shortening the duration of the trial. As the TITE-CRM generalized the CRM, the aaBCD generalizes the BCD by using weights in the dose-assignment algorithm that are adaptively updated during the trial and are a function of the amount of follow-up gathered on all currently enrolled subjects. Section 2.2 covers the details of the BCD, which then leads into our development and description of the aaBCD. Section 2.3 presents the results of simulations designed to examine the performance of the aaBCD and also compare the performance of the aaBCD to both the TITE-CRM and the ABCD. Section 2.4

contains concluding remarks.

2.2 Methods

2.2.1 The traditional BCD and the accelerated BCD

Suppose there are L doses under study, denoted as $d_1, \cdots, d_L$, among which we aim to find an MTD that is associated with a targeted probability of DLT, Γ. We plan to enroll a total of N patients. We assume that $\Gamma \leq 0.5$, although our methods can be applied to situations where $\Gamma > 0.5$ by replacing Γ with $1 - \Gamma$. We assign the first patient to one of the candidate doses, which, to promote patient safety, is usually the lowest dose, d_1. For all other patients, we use the following dose-assignment algorithm until all N patients have been enrolled. For subject j, $j = 2, \cdots, N$, if patient $j-1$ is assigned to dose d_l, $l = 1, \cdots, L$, and experiences a DLT, patient j is assigned to dose d_{l-1}. If $l = 1$, patient j is also assigned to the lowest candidate dose. If patient $j-1$ has no DLT by the end of the planned follow up period τ, we flip a biased coin with probability of heads equal to $p_h = \Gamma/1 - \Gamma$. If a head is observed, patient j is assigned to dose d_{l+1}, with the restriction that if $l = L$, patient j will be assigned to the highest candidate dose. If a tail is observed, patient j is assigned to dose d_l.

Durham and Flournoy (1995b) demonstrated that under certain conditions, the mode of the dose assignment distribution of the N patients can be used as a nonparametric estimate of the MTD. Suppose the true underlying probabilities of toxicity of the L candidate doses are $p_1, p_2, \cdots, p_L$. We denote the respective probabilities of

escalating, de-escalating or remaining at dose level l where $l = 2, \cdots, L-1$ are:

$$\begin{aligned} p_l^e &= (1-p_l)p_h \\ p_l^d &= p_l \\ p_l^r &= 1 - p_l^e - p_l^d \end{aligned}$$

The corresponding boundary probabilities for the lowest ($l = 1$) or highest ($l = L$) candidate doses are $p_1^d = p_L^e = 0$, $p_1^e = (1-p_1)p_h$, $p_L^d = p_L$, $p_1^r = 1-(1-p_1)p_h$, and $p_L^r = 1 - p_L$.

The doses are linked by the above transition probabilities which constitute a one-dimensional discrete random walk. Since all the doses are reachable from each other and can be consecutively assigned to patients, this random walk is irreducible and aperiodic. Furthermore, there is a finite expected number of patients needed for a dose to be assigned again from the last time it was assigned, making the random walk positive recurrent. Given these facts, this random walk has a unique stationary treatment distribution, which is also the limiting distribution $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_L)$, where $\pi_l, l = 1, \cdots, L$ is the limiting probability, as $n \to \infty$, of patient n being assigned to dose l.

Durham and Flournoy (1995b) showed that, when $(p_1^d, p_2^d, \cdots, p_L^d)$ is monotone decreasing and $(p_1^e, p_2^e, \cdots, p_L^e)$ is monotone increasing, the stationary treatment distribution $\boldsymbol{\pi}$ has a single mode at one of the candidate doses, say d_l^*, except when $p_{l^*}^e = p_{l^*}^d$, a situation for which the mode of $\boldsymbol{\pi}$ spans the interval $[d_{l-1}^*, d_l^*]$. If the underlying probabilities of DLT of the L candidate doses are equally spaced with $|d_l - d_{l-1}| = \Delta, l = 2, 3, \ldots L$, and if the probability of DLT is monotone increasing in dose with boundary conditions $p_2^d < p_1^e$ and $p_L^d > p_{L-1}^e$, the BCD will eventually select

a dose d_l^*, which is the empirical mode of the dose assignment distribution, such that $|\mu - d_l^*| < \Delta$, where μ is the true dose that has a probability of DLT equal to Γ.

As we have stated, the BCD requires complete follow-up of every patient and may result in long trials with a large proportion of time spent on waiting for results from previous patient(s). One of the approaches that eliminates the need for complete follow up is the accelerated Biased Coin Design (ABCD) of *Stylianou and Follmann* (2004). The ABCD is identical to the BCD when all patients are completely followed, but when the current patient has not completed their follow-up and a new patient is ready to be enrolled, the ABCD method assigns a dose to the new patient with the BCD algorithm using the outcome of the patient who has most recently completed their follow up. For example, consider a study in which the maximum follow up time is $\tau = 100$ days. Assume the first two patients have already been completely followed or have experienced a DLT, while the third patient has been followed for 60 days without experiencing DLT. At this point, we would assign the fourth patient to a dose based upon the result of the second patient, who is the last patient to have completed their follow-up.

When using the ABCD, we ignore the partial information collected from patients who have not completed their follow-up. Although the ABCD is very effective at reducing the total length of study, it loses precision by ignoring information from partially followed patients. Simulation studies in Section 2.3 demonstrate that the ABCD tends to pick a dose as the MTD that is lower than the true MTD, especially when the MTD is among one of the highest doses studied. As a result, we need a method that can eliminate the need for complete follow-up as well as one that can utilize the partial information from partially followed patients. This motivates a time-to-event version of the BCD, which we call the adaptive accelerated Biased

Coin Design (aaBCD).

2.2.2 The Adaptive Accelerated Biased Coin Design

We propose an adaptive weight function that assigns a small weight to the probability of escalation when the current patient has not completed their follow-up and also considers the outcomes of all patients assigned to the same dose level. Suppose dose l has been assigned to the current patient. Let m_l be the number of patients assigned to d_l with complete follow-up and without DLT, n_l be the number of patients assigned to dose d_l with DLT, and k_l be the number of patients assigned to d_l without DLT who have not yet completed their follow-up. Let $\bar{x}_l \leq \tau$ be the average amount of follow-up for the k_l patients with incomplete follow-up. When a new patient is ready for enrollment but the current patient is still being followed with no DLT yet observed, we flip a biased coin with the following probability of a head, which is also the probability of escalating the dose:

$$p_h^* = p_h \cdot \frac{m_l + k_l \cdot \frac{\bar{x}_l}{\tau}}{m_l + n_l + k_l} \leq p_h \tag{2.1}$$

The weight function $w_l = (m_l + k_l \cdot \bar{x}_l/\tau)/(m_l + n_l + k_l)$ computes the fraction of patients (assuming uniform DLTs over $[0, \tau]$) assigned to the current dose expected to not experience DLT by τ. By design, the weight makes it less likely to escalate a dose if more patients with DLTs are observed, i.e., when n_l is large. Conversely, the weight makes it more likely to escalate a dose if more patients without DLTs are observed, i.e., when m_l is large. Also, as $\bar{x}$ moves closer to τ, dose escalation becomes more likely and as $\bar{x}$ moves closer to 0, dose escalation becomes less likely.

We now study the limiting performance of the aaBCD. We use an indicator func-

tion I_l to indicate if there is any incomplete follow-up when a new patient can be enrolled, i.e., $I_l = 1$ if k_l=0 and $I_l = 0$ if $k_l \geq 1$. The probabilities of escalating, de-escalating or remaining at the current dose level l for the aaBCD are:

$$p_l^e = p_h(1-p_l)^{I_l} w_l^{1-I_l} \tag{2.2}$$

$$p_l^d = {p_l}^{I_l} I_l \tag{2.3}$$

$$p_l^r = 1 - p_l^e - p_l^d \tag{2.4}$$

Because the above transition probabilities depend on the presence of incomplete follow-up, the distributions of inter-arrival times and times of DLT are important in determining the transition probabilities in Equations (2.2) - (2.4).

Let $P_{D<A}$ be the probability that a DLT is observed before a new patient$'$s arrival, and $P_{A<\tau}$ be the probability that the inter-arrival time between a new patient and the most recently enrolled patient is less than τ. Then for any dose $2 \leq l \leq (L-1)$, the following one-step transition matrix applies:

$$\begin{aligned} p_l^e &= p_l(1-P_{D<A})p_h w_l + (1-p_l)p_h(1-P_{A<\tau}) + (1-p_l)p_h w_l P_{A<\tau} \\ p_l^d &= p_l P_{D<A} \\ p_l^r &= 1 - p_l^e - p_l^d \end{aligned}$$

The above transition probabilities depend on three quantities: (1) the weight w_l, (2) the distribution of DLT times, and (3) the distribution of inter-arrival times. These transition probabilities are consistent with and reduce to the transition probabilities of the regular BCD design when no partial follow-up exists, i.e., when $P_{D<A} = 1$ and $P_{A<\tau} = 0$. Also, the above limiting transition matrix is irreducible, aperiodic, and positive recurrent. Therefore, if w_l is fixed, for given $P_{D<A}$ and $P_{A<\tau}$, there is a

unique stationary vector of dose assignment probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_L)$ that is also the limiting distribution of the dose assignment vector and is the solution to the equations:

$$\pi_l p_l^d = \pi_{l-1} p_{l-1}^e \tag{2.5}$$

with boundary conditions $p_1^d = p_L^e = 0$ and constraint $\sum_{l=1}^{L} \pi_l = 1$, leading to the solution $\pi_1 = [1 + \sum_{l=2}^{L} \prod_{i=2}^{l} p_{i-1}^e / p_i^d]^{-1}$ and $\pi_l = \pi_{l-1} p_{l-1}^e / p_l^d$, where $l = 2, \cdots, L$.

The above limiting distribution does not necessarily target the MTD as the mode of $\boldsymbol{\pi}$ when $P_{D<A} \neq 1$ or $P_{A<\tau} \neq 0$, as it depends on the maximum follow up time τ, the distribution of patient inter-arrival times, and the distribution of DLT times. When τ is small relative to patient inter-arrival times, k_l is small relative to m_l and n_l. As a result, w_l will very often be 1 (when $k_l = 0$) or close to $m_l/(m_l + n_l)$ (when $k_l \neq 0$). These values for w_l cause p_h^* to be either the same as or slightly lower than p_h, leading to an MTD estimate from the aaBCD that is equal to or slightly lower than the mode of $\boldsymbol{\pi}$. Conversely, when τ is large relative to patient inter-arrival times, p_h^* is likely to be much less than p_h, thereby increasing the likelihood of the aaBCD to estimate the MTD at doses below the true MTD. However, despite the above facts, even with moderately large τ, the aaBCD still works well in terms of choosing the correct MTD since it mostly targets no more than one dose away from the MTD.

When the study has enrolled and completely followed all N subjects, we do not use the mode of the dose assignment distribution as our estimate of the MTD. Instead, we use an isotonic regression estimator which converges faster to the true MTD (*Stylianou and Flournoy*, 2002) than the mode of the dose assignment distribution. The isotonic estimator involves the following two steps. In the first step, we compute the observed proportions of patients with DLT $\hat{\boldsymbol{p}} = (\hat{p}_1, \cdots, \hat{p}_L)$ for candidate doses $d_1, \cdots, d_L$. If

the elements of $\hat{\boldsymbol{p}}$ are non-decreasing, we go to the second step. If the elements of $\hat{\boldsymbol{p}}$ are not non-decreasing, we scan the elements of $\hat{\boldsymbol{p}}$ sequentially from left to right. When we encounter an element of $\hat{\boldsymbol{p}}$ that is lower than the preceding element, we average those two elements and any others that are equal to or have been set equal to the preceding element. This process is also known as the pool adjacent violators algorithm (PAVA). We then go to the second step. In the second step, we examine the resulting vector $\hat{\boldsymbol{p}}$ from the first step and define the MTD as the dose d_l with the smallest value of $|\Gamma - \hat{p}_l|$. If more than one dose meets this criterion, we assign the MTD to the lowest (highest) of those doses if their corresponding values in $\hat{\boldsymbol{p}}$ all are greater (lower) than or equal to Γ. If two doses have probabilities equidistant from Γ, but on opposite sides of Γ, we conservatively assign the MTD to the lower of the two doses.

2.3 Simulation Studies

2.3.1 Description of Simulation Settings

Via simulation, we compared the operating characteristics of the aaBCD, ABCD and TITE-CRM in hypothetical trials of either four, five or six candidate doses, in which the candidate doses have DLT probabilities that are either equally spaced, with a targeted DLT rate of $\Gamma = 0.25$ or unequally spaced, with a targeted DLT rate of $\Gamma = 0.30$. For each of these settings, simulations were further stratified so each of the candidate doses was the true MTD. For example, when there were five candidate doses, we had five separate scenarios to examine in which the targeted DLT rate was located at the first, second, third, fourth, and fifth dose, respectively. To assess the impact of the distribution of DLT times, we ran all simulations assuming

DLTs occurred uniformly over the period of follow-up and then ran a second set of simulations where DLTs only occurred during the last quarter of follow up, termed late-onset DLTs.

For the TITE-CRM, the dose-toxicity relationship was modeled as $p_l = {d_l}^{exp(\beta)}$, where d_l is the initial estimate of toxicity probability ("skeleton") of dose level l and the model parameter β has a normal prior with mean 0 and variance 1. Due to the vast number of settings examined, a single "skeleton" could not be used across all settings. Thus, we chose to use the true DLT probabilities as the "skeletons" for TITE-CRM. As we discuss later, this approach will tend to overstate the performance of the TITE-CRM, as the true (unknown) DLT probabilities will deviate from the "skeleton" in any realistic study.

The total sample size for a trial was determined as six patients per candidate dose, i.e. 24 patients with four doses, 30 patients with five doses, and 36 patients with six doses. For a sample size of n patients, it was assumed that the duration of the study would be $M = 100n$, and the n (ordered) patient arrival times were drawn uniformly from the interval $[0, M]$, which is the same as assuming that inter-arrival times were not fixed, but rather varied uniformly with a mean of 100. We considered values for the length of follow-up $\tau \in \{0, 100, 200, 300\}$. Simulations were also performed with τ equal to 400, 500, and 600, but the results for those simulations are not shown as they led to results similar to those presented. The true probabilities of DLTs that are used in above simulations are listed in Table 2.1.

For each setting, we ran 2,000 simulations. Results of the 2,000 simulations were summarized so that the aaBCD, ABCD, and TITE-CRM could be compared with respect to: (1) the probability that the MTD was correctly identified, and (2) the average percent of patients assigned to each dose.

2.3.2 Simulation Results

2.3.2.1 Performance of aaBCD

Table 2.2 contains the probabilities of correctly identifying the MTD when the true vector of probabilities of DLTs are unequally spaced. In Table 2.2, we see that with uniform DLT times, the ability of aaBCD to correctly identify the MTD is fairly unaffected by the value of τ, the length of planned follow-up. For example, when there are six doses under study and the MTD is located at the second lowest dose, we see that the aaBCD correctly identifies the MTD anywhere from 45% to 50% of simulations depending on the value of τ. Thus, the performance of the aaBCD is fairly robust to the amount of incomplete follow-up observed in the trial when DLTs occur uniformly.

However, with late-onset DLTs, we see some distinct patterns. When the actual MTD is one of the highest doses studied, the ability of the aaBCD to correctly identify the MTD improves slightly when $\tau > 0$ as compared to when $\tau = 0$, i.e. no partial follow-up occurs. For example, when five doses are studied and the MTD is the highest dose, the aaBCD correctly identifies the MTD in 36% of simulations when $\tau = 0$ and 44% of simulations when $\tau = 200$. In contrast, when the true MTD is among one of the lowest doses studied, the ability of the aaBCD to correctly identify the MTD is best when $\tau = 0$ relative to when $\tau > 0$.

Table 2.3 contains the percentages of patients assigned to each dose when the true vector of probabilities of DLTs are unequally spaced. In Table 2.3, we see that in most settings, the percent of patients assigned to the true MTD is also relatively unaffected by the planned length of follow-up. One exception to this pattern is when the true MTD is located at the lowest dose studied. In that setting, we see that fewer

patients are assigned to the lowest dose when $\tau > 0$ relative to when $\tau = 0$. For example, when there are six doses studied and the true MTD is the lowest dose, we see that with uniform DLTs, 44% of patients are assigned to the lowest dose when $\tau = 0$, which decreases to 31% when $\tau = 300$. This differential is even greater with late-onset DLTs.

The results in Tables 2.2 and 2.3 exemplify how the use of partial follow-up can lead the aaBCD (and really any dose-finding algorithm) to be aggressive. Since our weights assume uniformly distributed DLTs, the aaBCD is more likely to escalate late in a patient′s follow-up because the weights indicate that a late-onset DLT is less likely than it really is. Thus, higher doses will tend to be preferred, and will be assigned to more patients. These properties are not necessarily a weakness of the aaBCD, as the liberal nature of the aaBCD will actually improve its performance when the MTD exists among the highest doses under study.

Tables 2.4 and 2.5 present results similar to those in Tables 2.2 and 2.3, respectively, but with equally spaced DLT probabilities, the setting in which the BCD performs optimally. In Table 2.4, we once again see that with uniform DLTs, the ability of the aaBCD to correctly identify the MTD is unaffected by the value of τ, but with late-onset DLTs, we see that the presence of incomplete follow-up ($\tau > 0$) can increase (decrease) the performance of the aaBCD when the true DLT is among the highest (lowest) doses studied. The results in Table 2.5 also lead to similar results as those reached from the results in Table 2.3.

2.3.2.2 Comparison of aaBCD to ABCD and TITE-CRM

In Table 2.2, we see that a vast number of values corresponding to the aaBCD and the TITE-CRM are boldfaced. This indicates that with both approaches, the

dose most often identified as the MTD at the end of the study is located at the true MTD. In contrast, many fewer of the values in Table 2.2 corresponding to the ABCD are boldfaced, because as τ becomes larger relative to the average inter-arrival time of patients, the ABCD becomes conservative and tends to identify the MTD at lower doses. This fact is emphasized in Table 2.3, where we see that when the ABCD is used, a majority of patients are assigned to the true MTD only when the true MTD is the lowest dose under study. Similar results are seen in Tables 2.4 and 2.5; thus, we determine that the ABCD is inferior to both the aaBCD and TITE-CRM.

In a head-to-head comparison of the aaBCD and the TITE-CRM, we see in Tables 2.2 and 2.3 that when the true MTD is among the highest doses studied, the TITE-CRM is more likely to correctly identify the MTD and assign more patients to that dose than is the aaBCD. However, the differential between the aaBCD and TITE-CRM largely vanishes when the true MTD lies among the middle of the range of doses studied. Moreover, we see in Tables 2.4 and 2.5 that the aaBCD can actually outperform the TITE-CRM when the true DLT probabilities are equally spaced, which is the setting where the optimality of the BCD was first proved.

As we stated earlier in Section 2.2.3.1, the "skeleton" used in our simulations for the TITE-CRM matched the true DLT probabilities. By doing so, we have contributed to an improved performance of TITE-CRM than what may be realized in actual trials. If we had used a "skeleton" in which the DLT probabilities were lower than the true DLT probabilities, the TITE-CRM could have been lead to prefer lower doses. Furthermore, the performance of the TITE-CRM is dependent upon the value of the prior variance on the model parameter β; differing values of the variance will impact the performance of the TITE-CRM. These facts also emphasize the inherent differences between the aaBCD and TITE-CRM, namely that the TITE-CRM is a

Bayesian model-based approach to dose-finding, whereas the aaBCD can be viewed as a non-parametric, model-free approach to dose-finding. As such, the aaBCD is straightforward to implement and execute, as it does not require specification of a model, the "skeleton" of prior DLT rates, the prior distribution of the model parameter, nor sophisticated statistical software to do the Bayesian calculations necessary for each dose assignment in the trial.

2.4 Conclusion and Discussion

By incorporating a weight function into the probability of escalation for the BCD, the aaBCD eliminates the need for complete follow-up of every patient, thereby shortening the duration of the entire trial. At the same time, the aaBCD is able to maintain a high probability of correctly targeting the MTD, as well as assign a larger percentage of patients to the actual MTD, relative to when no partial follow-up exists. In our simulations, we have seen that the aaBCD performs better than the ABCD, which tends to locate the MTD at lower doses as the length of follow-up increases. The aaBCD has performance similar to that of the TITE-CRM when the true MTD is near the middle of the range of doses studied and actually slightly outperforms the TITE-CRM when the true DLT probabilities are equally spaced.

Like the traditional form of the TITE-CRM, our weight function is based upon an assumption of uniformly distributed DLT times. However, the aaBCD can be easily generalized to accommodate early- or late-onset DLTs as needed. We also considered an alternate weight function $w_l = [(m_l + k_l\bar{x}_l/\tau)/(m_l + n_l + k_l)]^{k_l}$. The appeal of this weight function was that when $k_l = 0$, the probability of escalation reduced to that of the BCD. However, we found that as k_l increased away from zero, the probability of

escalation decreased too quickly and make escalation difficult, even when warranted. Another weight function we considered was $w_l = [m_l/(m_l + n_l + k_l)]^{k_l}$ which preferred lower doses more than the uniform weight function proposed in Section 2.2.2 when k_l is small. However, in simulations, we found this weight function tended to be almost as conservative as the ABCD when the target dose was among the highest candidates.

Future work can be emphasized on examining the applicability of the aaBCD to two bivariate settings. The first setting involves the use of the aaBCD in Phase I studies of combinations of two agents, which are growing in popularity in oncology as single-agent therapies for cancer are proving less successful than desired but the toxicity profiles of combined agents are unknown. Here the difficulty lies in determining how dose assignments should occur during the study, as each agent could be escalated or de-escalated simultaneously. The second setting involves the incorporation of efficacy into the dose assignment algorithm, so that the dose selected at the end of the study is not only non-toxic, but also appears to be therapeutic. Here the complication lies in how one computes the probability of escalation, as the decision to escalate must reflect whether or not the current patient has a positive response, in terms of efficacy, to the dose assigned.

Table 2.1: True DLT probabilities in scenarios examined in simulations. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.)

		Equally Spaced Probs. of DLT						Unequally Spaced Probs. of DLT					
	Target	1st	2nd	3rd	4th	5th	6th	1st	2nd	3rd	4th	5th	6th
4 Doses	1st	0.30	0.35	0.40	0.45	–	–	0.25	0.34	0.44	0.55	–	–
	2nd	0.25	0.30	0.35	0.40	–	–	0.17	0.25	0.33	0.45	–	–
	3rd	0.20	0.25	0.30	0.35	–	–	0.08	0.16	0.25	0.35	–	–
	4th	0.15	0.20	0.25	0.30	–	–	0.01	0.08	0.15	0.25	–	–
5 Doses	1st	0.30	0.35	0.40	0.45	0.50	–	0.25	0.38	0.48	0.55	0.60	–
	2nd	0.25	0.30	0.35	0.40	0.45	–	0.10	0.25	0.31	0.39	0.44	–
	3rd	0.20	0.25	0.30	0.35	0.40	–	0.03	0.16	0.25	0.31	0.35	–
	4th	0.15	0.20	0.25	0.30	0.35	–	0.01	0.11	0.19	0.25	0.29	–
	5th	0.10	0.15	0.20	0.25	0.30	–	0.01	0.09	0.16	0.21	0.35	–
6 Doses	1st	0.30	0.35	0.40	0.45	0.50	0.55	0.25	0.31	0.38	0.46	0.55	0.65
	2nd	0.25	0.30	0.35	0.40	0.45	0.50	0.14	0.25	0.36	0.45	0.55	0.64
	3rd	0.20	0.25	0.30	0.35	0.40	0.45	0.07	0.16	0.25	0.34	0.44	0.53
	4th	0.15	0.20	0.25	0.30	0.35	0.40	0.05	0.10	0.17	0.25	0.33	0.42
	5th	0.10	0.15	0.20	0.25	0.30	0.35	0.03	0.08	0.13	0.19	0.25	0.38
	6th	0.05	0.10	0.15	0.20	0.25	0.30	0.01	0.05	0.09	0.14	0.19	0.25

Table 2.2: Proportion of 2,000 simulations in which each dose was identified as the MTD when the doses have unequally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose that was selected most often on average. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.)

	DLT		4 Candidates target on				5 Candidates target on					6 Candidates target on					
Method	Occurrence	τ	4th	3rd	2nd	1st	5th	4th	3rd	2nd	1st	6th	5th	4th	3rd	2nd	1st
aaBCD	Uniform	0	**54**	**40**	**42**	**65**	**36**	**32**	**36**	**48**	**72**	**39**	30	**35**	**42**	**50**	**58**
		100	**52**	**41**	**41**	**60**	**37**	**34**	**36**	**46**	**71**	**42**	**32**	**35**	**42**	**48**	**53**
		200	**52**	**42**	**41**	**58**	**36**	**32**	**37**	**43**	**66**	**40**	**32**	**35**	**40**	**48**	**51**
		300	**50**	**45**	**42**	**61**	**33**	**32**	**38**	**45**	**64**	**36**	**32**	**37**	**41**	**45**	**50**
ABCD		0	**53**	**40**	**42**	**65**	**35**	29	**38**	**49**	**75**	**41**	30	**37**	**40**	**51**	**58**
		100	37	36	**43**	**71**	17	27	39	**51**	**74**	20	21	29	**40**	**52**	**64**
		200	24	35	**44**	**72**	8	21	36	**53**	**76**	8	12	25	38	**52**	**67**
		300	16	28	**43**	**71**	5	16	37	**54**	**77**	4	8	21	39	**49**	**68**
TITE-CRM		0	**73**	**43**	**33**	**69**	**61**	27	**36**	**52**	**79**	**67**	**36**	**43**	**48**	**53**	**67**
		100	**71**	**43**	**35**	**74**	**59**	26	**36**	**50**	**81**	**64**	**36**	**42**	**48**	**52**	**69**
		200	**67**	**41**	**36**	**72**	**54**	25	**37**	**52**	**81**	**62**	**35**	**41**	**49**	**53**	**70**
		300	**67**	**40**	**35**	**76**	**54**	24	**36**	**51**	**82**	**61**	**37**	**40**	**46**	**49**	**71**
aaBCD	Late Onset	0	**54**	**41**	**40**	**64**	**36**	29	**37**	**49**	**73**	**41**	**32**	**35**	**42**	**49**	**58**
		100	**58**	**43**	**38**	**57**	**42**	**33**	**37**	**42**	**63**	**45**	**34**	**35**	**40**	**45**	**45**
		200	**57**	**43**	**37**	**53**	**44**	**31**	**34**	**37**	**56**	**47**	**35**	**35**	**38**	**37**	**43**
		300	**55**	**42**	**38**	**51**	**39**	**32**	**36**	**39**	**56**	**45**	**34**	**34**	**38**	**40**	**41**
ABCD		0	**53**	**42**	**42**	**66**	**35**	29	**38**	**48**	**73**	**40**	31	**35**	**43**	**48**	**57**
		100	40	37	**42**	**66**	22	26	**40**	**52**	**76**	23	22	31	**43**	**51**	**62**
		200	26	35	**43**	**68**	14	24	38	**53**	**75**	11	16	31	**40**	**52**	**64**
		300	20	34	**44**	**68**	7	20	37	**54**	**74**	5	12	24	**41**	**52**	**64**
TITE-CRM		0	**74**	**41**	**35**	**70**	**60**	25	**38**	**52**	**79**	**65**	**38**	**42**	**47**	**53**	**67**
		100	**70**	**41**	34	**73**	**57**	26	**38**	**51**	**80**	**65**	**37**	**43**	**47**	**53**	**70**
		200	**67**	**41**	35	**75**	**56**	27	**37**	**52**	**82**	**65**	**38**	**42**	**49**	**51**	**71**
		300	**66**	**42**	34	**75**	**52**	25	**36**	**53**	**83**	**63**	**38**	**43**	**44**	**50**	**71**

Table 2.3: Average percentage of patients assigned to each dose in 2,000 simulations when the doses have unequally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose most often assigned to patients. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.)

			4 Candidates target on				5 Candidates target on					6 Candidates target on					
Method	DLT Occurrence	τ	4th	3rd	2nd	1st	5th	4th	3rd	2nd	1st	6th	5th	4th	3rd	2nd	1st
aaBCD	Uniform	0	28	25	32	**52**	18	20	24	31	**53**	18	17	21	25	32	**44**
		100	28	26	32	**46**	20	20	24	**30**	**46**	**20**	18	20	25	**31**	**38**
		200	**27**	25	32	**42**	19	19	23	**29**	**40**	18	17	19	24	**30**	**33**
		300	24	25	31	**42**	17	18	24	**30**	37	16	15	19	24	**28**	**31**
ABCD		0	27	26	33	**51**	18	20	24	31	**54**	18	17	20	24	33	**45**
		100	12	16	30	**67**	6	11	18	32	**66**	5	7	12	19	31	**61**
		200	5	11	25	**75**	2	5	13	29	**75**	1	3	7	13	28	**71**
		300	3	7	22	**80**	1	3	10	26	**80**	1	1	4	11	25	**78**
TITE-CRM		0	**60**	**32**	26	**64**	**51**	18.6	27	**39**	**71**	**54**	**25**	**30**	**33**	**38**	**62**
		100	**57**	**30**	27	**70**	**49**	18	27	**38**	**74**	**51**	**25**	**29**	**33**	**38**	**65**
		200	**54**	**30**	26	**71**	**44**	18	27	**38**	**76**	**50**	**24**	**29**	**33**	**38**	**68**
		300	**52**	**29**	26	**73**	**43**	18	26	**38**	**80**	**47**	**25**	**28**	**32**	36	**70**
aaBCD	Late Onset	0	28	25.4	32	**52**	18	19	24	31	**54**	19	18	2	25	32	**45**
		100	**34**	**27**	**29**	**35**	**27**	21	22	**25**	**34**	**26**	19	**20**	**22**	**25**	26
		200	**35**	25	**26**	**29**	**29**	19	20	**22**	25	**27**	17	17	**19**	20	20
		300	**33**	23	**26**	27	**25**	18	21	**22**	23	**24**	15	17	**18**	19	19
ABCD		0	27	26	33	**51**	18	19	24	31	**54**	19	17	20	25	33	**44**
		100	15	18	31	**59**	8	12	20	32	**61**	7	9	14	21	32	**53**
		200	15	18	31	**59**	8	12	20	32	**61**	7	9	14	21	32	**53**
		300	7	14	28	**67**	3	8	16	31	**66**	2	4	10	17	31	**60**
TITE-CRM		0	**59**	**31**	26	**65**	**49**	18	27	**38**	**71**	**55**	26	**28**	**33**	**38**	**62**
		100	**57**	**30**	26	**68**	**46**	18	27	**38**	**74**	**52**	25	**29**	**33**	**38**	**65**
		200	**54**	**30**	27	**72**	**46**	19	26	**39**	**77**	**51**	25	**29**	**34**	**37**	**67**
		300	**52**	29	26	**74**	**42**	18	26	**39**	**79**	**49**	**25**	**29**	**32**	36	**70**

Table 2.4: Proportion of 2,000 simulations in which each dose was identified as the MTD when the doses have equally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose that was selected most often. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.)

			4 Candidates target on				5 Candidates target on					6 Candidates target on					
Method	DLT Occurrence	τ	4th	3rd	2nd	1st	5th	4th	3rd	2nd	1st	6th	5th	4th	3rd	2nd	1st
aaBCD	Uniform	0	**40**	29	**33**	**50**	**37**	27	26	**32**	**51**	**37**	25	**26**	**28**	**32**	**50**
		100	**40**	**30**	**32**	**49**	**40**	**29**	**28**	29	**48**	**40**	**29**	26	**28**	**32**	**46**
		200	**40**	**30**	**32**	**47**	**38**	27	**28**	**29**	**46**	**41**	**27**	**26**	**27**	**31**	**43**
		300	**40**	**30**	**32**	**46**	**38**	**28**	**29**	**29**	**44**	**39**	26	**27**	**27**	**29**	**43**
ABCD		0	**39**	29	**34**	**52**	**37**	27	28	**32**	**52**	**39**	**27**	**26**	**28**	**31**	**52**
		100	26	26	34	**58**	21	20	28	32	**59**	18	17	22	29	34	**58**
		200	14	24	35	**63**	10	14	26	35	**64**	8	11	18	27	38	**63**
		300	10	20	37	**65**	5	11	22	35	**68**	4	7	14	25	36	**67**
TITE-CRM		0	**58**	22	22	**62**	**57**	24	**23**	24	**65**	**63**	25	27	**27**	26	**64**
		100	**58**	22	21	**64**	**60**	25	**24**	23	**65**	**62**	26	26	**26**	26	**64**
		200	**55**	23	23	**64**	**57**	22	22	24	**67**	**59**	26	26	**26**	25	**67**
		300	**54**	21	21	**67**	**54**	26	**24**	23	**69**	**55**	26	27	**26**	26	**69**
aaBCD	Late Onset	0	**39**	29	**33**	**52**	**36**	26	27	31	**51**	**38**	26	**26**	27	**31**	**53**
		100	**47**	**32**	**28**	**42**	**48**	**30**	**26**	**29**	**41**	**47**	**29**	**26**	**26**	**29**	**38**
		200	**50**	**31**	**28**	**38**	**48**	29	**29**	**27**	**36**	**49**	**29**	**25**	**25**	**26**	**35**
		300	**47**	30	**28**	**38**	**47**	29	26	**27**	**35**	**50**	29	**25**	**23**	**24**	**35**
ABCD		0	37	30	32	**52**	**39**	**29**	26	31	**49**	**38**	**28**	26	**27**	**32**	**51**
		100	29	27	34	**55**	25	23	28	32	**55**	22	19	24	27	**35**	**54**
		200	21	26	37	**57**	14	20	27	33	**57**	14	15	22	28	33	**60**
		300	14	25	35	**62**	10	17	29	33	**62**	6	1	19	29	**37**	**60**
TITE-CRM		0	**61**	23	22	**61**	**59**	24	**25**	24	**63**	**63**	24	**26**	**26**	26	**65**
		100	**58**	21	21	**62**	**59**	25	**25**	24	**67**	**60**	26	**26**	**26**	26	**66**
		200	**55**	24	22	**65**	**59**	24	**24**	23	**64**	**58**	25	**25**	**26**	26	**65**
		300	**52**	22	20	**70**	**57**	25	23	23	**68**	**55**	24	**26**	24	24	**69**

Table 2.5: Average percentage of patients assigned to each dose in 2,000 simulations when the doses have equally spaced probabilities of DLT. Boldface indicates the mode, i.e. the dose most often assigned to patients. τ is the length of follow-up for each patient. Rankings of doses are from lowest to highest.(i.e., 1st means the lowest dose, and 2nd means the second lowest dose, etc.)

			4 Candidates target on				5 Candidates target on					6 Candidates target on					
Method	DLT Occurrence	τ	4th	3rd	2nd	1st	5th	4th	3rd	2nd	1st	6th	5th	4th	3rd	2nd	1st
aaBCD	Uniform	0	24	24	30	**44**	20	18	21	28	**41**	19	17	18	21	28	**40**
		100	**26**	23	29	**40**	22	19	22	27	**36**	**21**	17	18	21	26	**33**
		200	25	23	29	**37**	21	19	22	26	**33**	**20**	16	18	21	24	**29**
		300	24	23	29	**35**	21	18	22	26	**31**	19	15	18	20	24	**27**
ABCD		0	24	23	30	**44**	20	19	22	28	**43**	19	17	18	211	28	**41**
		100	10	15	29	**61**	7	9	15	29	**59**	5	7	10	16	29	**58**
		200	4	9	25	**72**	2	4	10	25	**70**	2	3	6	11	26	**69**
		300	2	6	20	**78**	1	21	17	211	**77**	1	1	3	8	22	**77**
TITE-CRM		0	**49**	17	18	**61**	**49**	18	18	19	**61**	**51**	18	**19**	19	19	**62**
		100	**47**	17	17	**64**	**48**	17	18	18	**65**	**49**	19	19	19	20	**64**
		200	**43**	16	17	**67**	**45**	17	17	18	**66**	**45**	18	19	18	19	**67**
		300	**42**	16	17	**69**	**43**	16	17	18	**70**	**43**	17	18	19	19	**68**
aaBCD	Late Onset	0	23	23	30	**44**	20	19	21	28	**42**	19	17	18	21	27	**41**
		100	**35**	25	25	**29**	**32**	21	21	**23**	**26**	**30**	19	17	**18**	20	22
		200	**39**	22	23	24	**36**	19	19	**19**	20	**33**	17	15	16	16	17
		300	**38**	21	22	23	**34**	17	18	**18**	19	**32**	15	15	15	15	16
ABCD		0	23	23	30	**44**	21	19	22	28	**42**	19	17	18	21	10	**40**
		100	13	18	30	**54**	10	12	19	29	**52**	8	9	13	19	30	**49**
		200	7	13	29	**62**	4	8	15	29	**59**	3	5	9	15	29	**58**
		300	3	10	26	**68**	2	5	12	27	**65**	1	3	6	1	28	**63**
TITE-CRM		0	**49**	17	18	**40**	**50**	17	19	18	**61**	**51**	18	**19**	19	19	**62**
		100	**46**	17	17	**42**	**46**	17	18	18	**65**	**48**	18	18	19	20	**65**
		200	**42**	17	17	**44**	**45**	17	18	18	**65**	**45**	18	18	19	19	**67**
		300	**39**	16	17	**47**	**43**	17	17	18	**69**	**43**	17	18	18	19	**68**

CHAPTER III

A Generalized Continual Reassessment Method for Two-Agent Phase I Trials

3.1 Introduction

The primary goal of traditional one agent phase I oncology trials is identifying the maximum tolerated dose (MTD), defined as the largest possible dose that maintains the probability of a dose-limiting toxicity (DLT) that is closest to a pre-specified target rate, $\Gamma \in (0, 1)$. The most widely used one agent parametric model, the CRM assumes that the probability of DLT is related to dose d through a simple parametric function $F(\beta; d)$ that is monotonic in dose, with monotonicity enforced by requiring $\beta > 0$. The two most commonly-used forms for $F(\beta; d)$ are the logistic model $\text{logit}[F(\beta; d)] = a + \beta d$ with known value a, where $\text{logit}(p) = \log(p) - \log(1-p)$, and the power model $F(\beta; d) = d^{\beta}$. Many recent Phase I trials have begun to study the combined toxicity of two agents, with multiple doses of both agents being studied. For example, *Rowinsky et al.* (1996) conducted a trial where cisplatin and topotecan with five and two dose levels, respectively, were simultaneously administered to study the combined toxicity of the two drugs. *Yuan and Yin* (2008) also present a study for

the combination of bortezomib with chemotherapy that examined eight dose levels for each agent. In such studies, identifying a maximum tolerated combination (MTC) cannot be done using single-agent designs. This is because even with an assumption of monotonicity of DLT rates between dose levels in one agent when the other agent is fixed, we still do not have a complete ordering of the probabilities of DLT for all possible dose combinations with the two agents. For instance, in a simple case of two doses each of Agents A and B, we have no knowledge regarding the order of the DLT rates of the combinations of the high dose of A and the low dose of B versus the low dose of A and the high dose of B.

Numerous statistical approaches exist for the design of two-agent Phase I trials. *Thall et al.* (2003) propose a design in which the DLT rate of a combination is modeled as a function of the doses of the two agents via a logistic model. Identification of the MTC is based on a contour map, in which combinations on the same contour have the same DLT rate. *Wang and Ivanova* (2005) use an alternate parametric model and define a "start-up" stage when little data are available for parameter estimation. After collection of enough data for a sufficient toxicity profile, the dose combinations deemed "acceptable" in the "start-up" stage are more fully examined using the parametric model and Bayesian methods for estimation. *Yuan and Yin* (2008) define a search range of combinations by first running a one-dimensional trial and later updating it, then eliminating the doses that lay outside of the search range based on the partial ordering of the DLT rates, thereby largely reducing the two-dimensional dose-finding space. Additional adaptive, parametric approaches have been proposed by *Yin and Yuan* (2009a), *Yin and Yuan* (2009b) and *Braun and Wang* (2010). Most recently, using the methods of partial ordering first suggested in the single-agent setting by *Conaway et al.* (2004), *Wages et al.* (2011a) and *Wages et al.* (2011b) develop designs

for two-agent Phase I trials based upon an examination of all possible orders of combinations, when the number of combinations can be feasibly enumerated, or a subset of all possible orders when enumeration of all possible orders is infeasible.

We propose an alternative design of two-agent Phase I trials that is a straightforward generalization of the CRM, which we denote gCRM. We visualize the combinations of m doses of Agent A and n doses of Agent B as a grid of m columns and n rows; see Figure 3.1(a) for an example when $m = n = 4$. We then adopt the CRM separately for each row in the grid, with each CRM specific to a single dose of Agent B in combination with doses of Agent A. We include a parameter β that is shared among all the row-wise models, while each model contains a second parameter (usually fixed in the one agent CRM) that is specific to the dose of Agent B being studied. These row-specific parameters are given a prior distribution that reflects their correlation and allows for "borrowing of strength" among all combinations being studied. In Section 3.2, we explicitly develop our model and describe how to determine hyperparameter values for the prior distributions of the model parameters, as well as rescaled dose values for one of the agents, using information elicited from investigators. We also explicitly outline the conduct of an actual trial using our methods. We examine the operating characteristics of our approach via simulations and compare our approach to other existing approaches in Section 3.3 . We present concluding remarks in Section 3.4.

3.2 Methods

3.2.1 Model Description

As stated in Section 3.1, the logistic model and the power model are commonly used as the dose-toxicity function $F(\beta; d)$ for the traditional CRM. We now generalize the CRM to a Phase I trial of two agents specific to the logistic model; a short description of generalization for the power model appears at the end of Section 3.2.2. We have a trial of two agents, in which there are m doses, $a_1 < a_2 < \ldots < a_m$, of Agent A, and n doses, $b_1 < b_2 < \ldots < b_n$, of Agent B. We let (j, k) represent the combination of Agent A at dose $a_j, j = 1, 2, \ldots, m$ and Agent B at dose $b_k, k = 1, 2, \ldots, n$. We let p_{jk} denote the true DLT probability for (j, k) and define the MTC, (j^*, k^*), as the combination with a DLT rate closest to the target rate Γ. For each value of k (each dose of Agent B), we adopt a traditional logistic regression model $\text{logit}(p_{jk}) = \alpha_k + \beta a_j$ describing how the probability of DLT varies with each dose of Agent A combined with the given dose of Agent B. Note that the intercept is now a parameter specific to the dose of Agent B under study, rather than a fixed value traditionally used in the CRM; see Figure 3.1(b) for a visual representation of this idea. In more general applications, this model is known as a proportional odds logistic regression model. We attempt to "join" the row-wise CRM models through a joint prior distribution that forces correlation among their intercept parameters. We note that like the traditional CRM, the dose values $a_1, a_2, \ldots, a_m$ of Agent A are not actual clinical dose values, but are rescaled dose values selected to improve model fit. However, there is no need to compute rescaled dose values for Agent B because the doses of Agent B do not contribute directly to our model.

We assume $\log(\beta)$ has a normal distribution with mean θ and variance σ^2, denoted

$\log(\beta) \sim \mathcal{N}(\theta, \sigma^2)$, and the intercept for the model at the lowest dose of Agent B, $\alpha_1 \sim \mathcal{N}(\mu, \sigma^2)$. Instead of directly modeling the joint distribution of the intercepts, we focus on the difference between intercepts corresponding to adjacent doses of Agent B. Specifically, we assume $\Delta_k = \alpha_k - \alpha_{k-1} \sim \mathcal{N}(\delta_k, \sigma^2), k = 2, 3, \ldots, n$. We note that our use of a normal distribution for the differences does allow for the possibility of non-increasing DLT rates among doses of Agent B for a fixed dose of Agent A. However, such a possibility is minimized if the value of σ^2 is kept sufficiently small so as to force a majority of the prior distribution above zero.

Let N_{jk} be the number of patients assigned to (j, k), and Y_{jk} be the number of the N_{jk} patients with DLT. Then the posterior distribution of $(\boldsymbol{\alpha}, \beta)$ is

$$f(\boldsymbol{\alpha}, \beta | \mathbf{Y}, \mathbf{N}) = \frac{\prod_{i=1}^{m} \prod_{j=1}^{n} f_{ij}(Y_{ij} | \boldsymbol{\alpha}, \beta, N_{ij}) g(\boldsymbol{\alpha}) h(\beta)}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{i=1}^{m} \prod_{j=1}^{n} f_{ij}(Y_{ij} | \boldsymbol{\alpha}, \beta, N_{ij}) g(\boldsymbol{\alpha}) h(\beta) d(\boldsymbol{\alpha}) d(\beta)}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_n)$,$\mathbf{Y} = \{Y_{ij} : i = 1, 2, \cdots, m, j = 1, 2, \cdots, n\}$, $\mathbf{N} = \{N_{ij} : i = 1, 2, \cdots, m, j = 1, 2, \cdots, n\}$, $\prod_{i=1}^{m} \prod_{j=1}^{n} f_{ij}(Y_{ij} | \boldsymbol{\alpha}, \beta, N_{ij}) = \binom{N_{ij}}{Y_{ij}} p_{i,j}^{Y_{ij}} (1 - p_{i,j})^{(N_{ij} - Y_{ij})}$, $g(\boldsymbol{\alpha})$ is the multivariate normal priors for $\boldsymbol{\alpha}$ with parameters μ and σ^2, and $h(\beta)$ is the log normal density for β with parameters θ and σ^2.

Since a closed-form expression for the posterior distribution is not available, based upon the the accumulated data $\boldsymbol{Y} = \{Y_{jk} : j = 1, 2, \ldots, m, k = 1, 2, \ldots, n\}$ and $\boldsymbol{N} = \{N_{jk} : j = 1, 2, \ldots, m, k = 1, 2, \ldots, n\}$, we can use Gibbs sampling to generate a sample of observations from the joint posterior distribution of $(\boldsymbol{\Delta}, \alpha_1, \beta)$, where $\boldsymbol{\Delta} = \{\Delta_2, \Delta_3, \ldots, \Delta_n\}$, which, in turn, gives us a sample from the posterior distribution of $(\boldsymbol{\alpha}, \beta) = \{\alpha_1, \alpha_2, \ldots, \alpha_n, \beta\}$. From this posterior sample, we can compute a posterior estimate for the probability of DLT for combination (j, k) as $\text{logit}(\tilde{p}_{jk}) = \tilde{\alpha}_k + \tilde{\beta} a_j$, in which $\tilde{\beta}$ and $\tilde{\alpha}_k$ are the respective posterior means of β and α_k.

3.2.2 Deriving Hyperparameter Values and Rescaled Doses

Before the trial can begin, investigators must supply the values p^*_{j1} and p^*_{1k}, which are the *a priori* estimates for p_{j1} and p_{1k}, the DLT rates for combinations that use the lowest dose of either or both of the agents, i.e the first column and row of Figure 3.1(a). This leads to $m+n-1$ values from which we seek to determine values for the $n+2$ hyperparameter values $\theta, \mu, \boldsymbol{\delta} = \{\delta_2, \delta_3, \ldots, \delta_n\}$, and σ^2, as well as the m rescaled dose values for Agent A. Due to this overparameterization relative to the amount of elicited information, we choose to treat μ, θ and σ^2 as "tuning parameters" whose values will be determined via simulation, with the remaining values based directly upon the elicited information. Specifically, if we assume from our logistic model that $\text{logit}(p^*_{j1}) = E(\alpha_1) + a_j E[\beta]$, where $E(\cdot)$ denotes prior expectation, we have the rescaled dose value $a_j \approx [\text{logit}(p^*_{j1}) - \mu]/\exp(\theta)$. Since δ_k is approximately the average difference between the logits of the DLT rates of doses b_k and $b_{k-1}, k = 2, 3, \ldots, n$ when each is combined with the same dose of Agent A, we set $\delta_k = \text{logit}(p^*_{1k}) - \text{logit}(p^*_{1,k-1})$.

An appropriate value for σ^2 requires that, in the starting stage of a trial when only a few patients are entered, the priors are informative enough to accommodate for the relative lack of data. In contrast, when more patients are enrolled later in the study, σ^2 should be "large enough" so that the data are able to dominate the priors. Also, given that the DLT rate for combination $(1, 1)$ is typically quite small, we would expect a rather large negative value to be appropriate for α_1. Based upon an extensive set of small simulation studies over a grid of candidate values for μ, θ, and σ^2, we have found that our design is able to identify the MTC well when using values of $\sigma^2 \in [0.3, 0.6]$, $\mu \in [-8.5, -7.5]$, and $\theta \in [4.5, 5.5]$.

We note that although the methods described here and in Section 3.3 are specific

to the use of a logistic model, our idea is easily generalized to any one-parameter model used with the CRM. For example, a simple transformation of the power model gives us $\log[-\log F(\beta; d)] = \alpha + \log[-\log(d)]$, where $\alpha = \log(\beta)$, which is a standard logistic model with complementary log-log link and slope parameter fixed at 1.0. Application of our methods to this complementary log-log model is straightforward. Adopting the previous parameterization used for the logistic model, for each dose k of Agent B, we have $\log[-\log(p_{jk})] = \alpha_k + \beta \log[-\log(a_j)]$, in which the slope parameter is now allowed to vary from the fixed value of 1.0, if necessary. We once again define $\Delta_k = \alpha_k - \alpha_{k-1}, k = 2, 3, \ldots, n$ and adopt the same form for the prior distributions of α_1, β, and $\mathbf{\Delta}$. However, appropriate values for the hyperparameters μ, θ, and σ^2 will necessarily change and will require the same sort of grid search via simulation to determine what those values might be. For the scenarios presented in Section 3.3, we found that $\theta = 3$, $\mu = 5.5$ and $\sigma^2 = 0.3$ led to satisfactory operating characteristics (not presented). We use the elicited information to set $\delta_k = \log[-\log(p^*_{1k})] - \log[-\log(p^*_{1,k-1})]$ and the rescaled dose values a_j are determined from the equation $\log[-\log(p^*_{j1})] \approx \mu + \exp(\theta)\log[-\log(a_j)]$. The simulation results for the power model gCRM are shown in Table 3.6.

From Table 3.6 we see that, the gCRM using the power CRM model maintains similar operating characteristics as the gCRM using the logistic model in terms of both the percentages of correct MTC identification, and the mean percentages of patients assigned to the recommended dose combinations. For example, using the gCRM with the power model, the percentages of the MTC selections whose DLT rates are within 10 points of Γ (or the percentage of selecting any dose combination as the MTC without terminating the trial for scenario D) for the seven scenarios are 95%, 93%, 66%, 3%, 77%, 87%, 91% respectively, which are comparable to their corresponding

counter parts of 90%, 95%, 86%, 5%, 74%, 84%, 91% from the gCRM with the logistic model; and the corresponding average percentages of patient assignments of the seven scenarios with the same criterion are 75%, 84%, 54%, 35%, 59%, 78%, 69%, which are similar to their respective counter parts of 77%, 84%, 75%, 40%, 59%, 81%, 75% from the gCRM using the logistic model. The gCRM using the power model is more likely to terminate a trial as compared to the gCRM using the logistic model. This is reflected in scenario C, where the targeted dose combination is one of the lowest dose combinations while the majority of the dose combinations are overly toxic in that, the gCRM using the power model terminates an additional of 20% of the simulations as compared to the gCRM using the logistic model; and correspondingly an additional of 21% of the patients are entered into a trial that uses the logistic model as compared to a trial using the power model.

3.2.3 Trial Conduct

Two outstanding issues remain for our design. First, like most Phase I trial designs, we restrict the set of possible assignments to each patient in the trial so as to limit exposure of patients to toxic combinations but also promote sufficient exploration of the combinations and increase the likelihood of correctly identifying the MTC. The idea of sufficient exploration is of greater import when studying combinations of two agents instead of a single agent because every additional dose level in an agent leads to a significant increase in the total number of dose combinations. To this end, the first patient is assigned to combination $(1, 1)$ and each future patient $i = 2, 3, \ldots N$ must be assigned to a combination that has doses of Agent A and Agent B that are within one dose (higher or lower) than the respective doses of Agent A and B assigned to patient $i - 1$. Note that this rule allows for the possibility of escalation

of both doses at the same time, which may seem aggressive, but has been shown to not lead to increased exposure to unsafe combinations (*Wang and Ivanova* (2005); *Braun and Wang* (2010)). We limit the amount of dose de-escalation to within one dose lower than the respective doses of Agent A and B because in two-agent trials, especially when few patients are entered into a study, a single DLT can be quite influential and result in dramatic dose de-escalation, which slows later dose escalation when the true MTC is among combinations including the highest dose of either agent.

Second, the study should be terminated if there is sufficient evidence that all of the combinations are overly toxic. Although a stopping rule could be based solely on the posterior distributions of the model parameters, we have found that such a rule leads to early termination too often. Instead, our design does not allow for early stopping until data have been collected on the third patient. At this point, the accumulated data from all patients, regardless of the combinations to which they were assigned, are pooled together to compute an exact 95% confidence interval for the probability of DLT. If the lower bound of this confidence interval is greater than the targeted DLT rate, Γ, the trial will be terminated.

Thus, the conduct of a Phase I trial for two agents using the gCRM is as follows:

(1) Elicit *a priori* estimates for the probabilities of DLT for combinations including the lowest dose of either agent;

(2) Determine appropriate values for the hyperparameters and rescaled dose values as described in Section 3.2.2;

(3) Assign the first patient to combination $(1, 1)$;

(4) For patient $i = 2, 3, \ldots N$, compute the posterior means $\tilde{p}_{jk}$ for each combination

(j,k). If we let (j_{i-1}, k_{i-1}) denote the combination assigned to patient $i-1$, we define the set of acceptable combinations as $S = \{(j,k) : \max(1, j_{i-1}-1) \leq j \leq \min(j_{i-1}+1, m), \max(1, k_{i-1}-1) \leq k \leq \min(k_{i-1}+1, n)\}$. The current estimate of the MTC, (j^*, k^*), is the combination in S whose value of $\tilde{p}_{jk}$ is closest to Γ;

(5) If $i \leq 3$, assign patient i to the current MTC;

(6) If $i \geq 4$, determine if the stopping rule above has been met. If so, terminate the trial; otherwise, assign patient i to the current MTC;

(7) If data have been collected on all N patients, use step (4) to compute the final estimate of the MTC.

3.3 Simulation Studies

3.3.1 Simulation results of the gCRM

We examine the operating characteristics of our design in seven scenarios, scenarios A-G, with $m = 4$ doses of Agent A and $n = 4$ doses of Agent B. Scenarios A-F were previously examined in *Braun and Wang* (2010). We have a target DLT rate of $\Gamma = 0.20$. The true DLT rates of the 16 combinations in each scenario are displayed in columns 3-6 of Table 3.1. In scenario A, combinations of lower doses of both agents have DLT rates less than Γ while the combinations of higher doses of both agents have DLT rates greater than Γ. Scenarios B and D reflect settings when all the combinations are safe and too toxic, respectively. Scenario C is a situation when only a few dose combinations are tolerable while all others are overly toxic. Scenario E is an example when the DLT rates grow slowly with doses of one agent but rapidly with doses of the other agent. Scenario F represents a situation when there is a jump in

the DLT rates of one agent, thus only the dose combinations associated with certain levels of that agent are safe. Scenario G reflects a situation when there are large deviations from the assumed proportional odds logistic model. It should be noted that none of true DLT rates fit our assumed proportional odds model.

Since we have 16 dose combinations for each scenario, we chose to use $N = 35$ as the maximum sample size of each trial in our simulations, which is approximately two patients per dose combination on average, and we consider that to be the minimal possible sample size for exploration of all dose combinations. We also used our design with increased sample sizes of $N = 40$ and $N = 50$, but we did not see a great difference in operating characteristics as compared to when $N = 35$, similar to the findings of *Braun and Wang* (2010).

We ran 2,000 simulations for each of the seven scenarios. We used the first row and first column of true DLT rates for scenario A as the elicited values supplied by the investigators. Thus, scenario A represents the situation when the elicited values equal the truth, while the remaining scenarios represent the more realistic setting when the elicited values do not equal the truth. These elicited values generated rescaled dose values of $a_1 = 0.0325, a_2 = 0.0374, a_3 = 0.0405$, and $a_4 = 0.0427$ and hyperparameters $\delta_2 = 0.98, \delta_3 = 0.54$, and $\delta_4 = 0.39$. We ran simulations using both the fixed hyperparameters and uniformly distributed priors on hyperparameters, but found there is no noticeable difference in the simulation results. Therefore, for simplicity reason, we recommend to use a fixed set of hyperparameters. When using a fixed set of hyperparameters, given the recommended hyperparameter ranges in Section 3.2.2, we simply chose $\mu = -8$, $\theta = 5$ and $\sigma^2 = 0.5$, while other combinations yield similar results.

The simulation results on scenarios A – G using uniform priors on the hyperpa-

rameters are presented in Table 3.5, and the simulation results on scenario A using fixed hyperparamters, but other combinations than the one discussed above are presented in Table 3.10. From both of the two tables we see that, the different settings of the hyperparameters certainly result in some differences on the simulation results. However, these differences are ignorable because the percentages of most of the quantities are no more than 2 percents′ different from setting to setting. This finding supports our decision that a fixed set of hyperparameters can be randomly selected from their respective domains.

In each simulation, we drew 10,000 samples from the posterior distributions with a burn-in of 2,000 samples. The samples were obtained via the "rjags" library in R. Among the 2,000 simulations in each scenario, we calculated the mean percentage that each dose combination was selected as the MTC at the end of the trial, the mean percentage of patients assigned to each dose combination, and the mean probability of DLT that each patient is exposed to based upon the their assignments in the 2,000 simulations. The simulation results for each scenario are summarized numerically in the last eight columns of Table 3.1 and visually in Figure 3.2.

As we summarize the results in Table 3.1, we note that with a sample size of $N = 35$ patients, it is nearly impossible to distinguish between DLT rates that are within 10 points of the targeted DLT rate. Thus, we define a combination to be acceptable for selection as the MTC if its true DLT rate is within this ten-point window. In scenario A, the gCRM selected a dose combination within the ten-point window as the MTC in 92% of simulations while assigning 81% of the patients to a dose combination within the ten-point window. Furthermore, in 46% of the simulations, the selected MTC is within a four-point window of Γ. In scenario B, all the dose combinations are safe and the gCRM selected the highest possible dose combination $(4, 4)$ in 62% of

simulations and selected the MTC at a combination within the ten-point window in 95% of simulations. Also, more than half of the patients are assigned to $(4, 4)$, which is a promising percentage considering the fact that we always assign the first patient to $(1, 1)$.

In scenario C, there are only four combinations in the ten-point window, and they are selected as the MTC in 86% of simulations and are assigned to 75% of patients. Moreover, combination $(2, 1)$, which has a DLT rate equal to Γ, is selected as the MTC in 39% of simulations. Due to the large number of combinations with DLT rates above Γ, 8% of simulations resulted in early trial termination. The remaining 6% of simulations selected a toxic combination as the MTC, although none selected a combination with a DLT rate above 0.4. In scenario D, all combinations are toxic, and the gCRM led to early termination in 95% of simulations, with combination $(1, 1)$ selected as the MTC in the remaining 5% of simulations.

In scenario E, the gCRM identified an MTC within the ten-point window in 74% of simulations and the identified MTC is within a two-point window in 42% of simulations. Since the elicited values from the investigator are based on the DLT rates of scenario A, we see that the gCRM is reluctant to explore higher doses of Agent B, since the gCRM assumes every escalation in Agent B will result in a bigger increase in the DLT rates than what the actual data suggest. 59% of the patients are assigned to the dose combinations within the ten-point window, and most of those assignments are within a two-point window. In scenarios F and G, whose DLT rates deviate greatly from the assumed proportional odds logistic model, we continued to see excellent operating characteristics for the gCRM.

Figure 3.2 presents the patterns of dose assignments in the seven scenarios. The y-axis is the average DLT rate of the 2,000 dose assignments of each patient and

the x-axis is the number of each patient (1 denotes the first patient, 2 denotes the second patient, etc). Note that the average DLT rate for each patient is only based upon trials in which that patient was enrolled, i.e. the mean dose for patient 30 would be based upon less than 2,000 simulations if some of the simulations had early termination of the trial. We see that the gCRM does an excellent job of converging to combinations with DLT rates equal to Γ, while in scenarios B and D, the gCRM tends to converge to combinations with the highest possible DLT rate and lowest possible DLT rate, respectively.

We also ran simulations on situations with four levels of Agent A and three levels of Agent B, five levels of Agent A and three levels of Agent B, as well as five levels of Agent A and four levels of Agent B. These results are presented in Table 3.7 – Table 3.9. From these three tables we can see that, in all scenarios except scenarios D1, D2 and D3, the MTC selections are mostly centered around the targeted dose combination whose DLT rate equals Γ, and the patient assignments also concentrates around the targeted dose combination. In scenario D1, D2 and D3, the gCRM also maintains high termination rate of at least 94%, enrolls no more than 40% of patients into a toxic trial, and the majority of the 40% patients are assigned to the lowest possible dose combination. It should be noted that, even for scenarios A3 – G3, where there are 20 dose combinations in each of the scenarios, the sample size used is still 35 patients, which is quite small as compared to the total number of dose combinations under study.

3.3.2 Comparison to Existing Approaches

Table 3.2 directly compares the operating characteristics of the gCRM with the methods of *Braun and Wang* (2010) in scenarios C, E, and F. Results for the other

scenarios are omitted as both methods were very similar in terms of both the percentage of simulations selected as the MTC for each dose combination, as well as the mean percentage of subjects assigned to each dose combination.

We see from Table 3.2 that the gCRM targets an MTC within the ten-point window more frequently than the method of *Braun and Wang* (2010): 86% versus 71% for scenario C; 74% versus 72% for scenario E; and 84% versus 67% for scenario F. The two approaches have comparable performance in terms of the mean percentages of patients assigned to a combination within the ten-point window. However, the gCRM assigns more patients than *Braun and Wang* (2010) at the combination with DLT probability exactly equal to Γ. For example, in scenario C, the combination with DLT rate exactly equal to Γ is assigned to 25% of patients with the gCRM and 21% with the method of *Braun and Wang* (2010), and in scenario F, the respective values are 12% and 6%. Correspondingly, the gCRM also assigns less patients to overly toxic dose combinations than *Braun and Wang* (2010).

We also compare the operating characteristics of the gCRM to the CRM for Partial Ordering (POCRM) of *Wages et al.* (2011a) in the six scenarios, scenarios 1-6, presented in their manuscript. Scenarios 1-4 correspond to scenarios 6, 3, 10 and 12 of *Yin and Yuan* (2009a), and scenarios 5 and 6 correspond to scenarios 1 and 2 of *Yin and Yuan* (2009b). Like *Yin and Yuan* (2009a) and *Yin and Yuan* (2009b), we used a sample size of $N = 60$ with the gCRM, while *Wages et al.* (2011a) used the average number of patients that were actually entered into the study, which will be less than 60 in settings where early termination occurred with *Yin and Yuan* (2009a) and *Yin and Yuan* (2009b).

The operating characteristics of the gCRM are based on 2,000 simulations in all six scenarios, and the hyperparamters of the prior distributions are the same as pre-

viously discussed in Section 3.1. In scenarios 1 and 2, we used the same elicited values as we used in Section 3.1. In scenarios 3 and 4, we used the elicited values $(p_{11}, p_{21}, p_{31}, p_{41}, p_{51}) = (0.04, 0.08, 0.12, 0.12, 0.20)$ and $(p_{11}, p_{12}, p_{13}) = (0.04, 0.10, 0.16)$, and in scenarios 5 and 6, we used $(p_{11}, p_{21}, p_{31}, p_{41}, p_{51}) = (0.04, 0.08, 0.12, 0.16, 0.20)$ and $(p_{11}, p_{12}, p_{13}, p_{14}) = (0.04, 0.10, 0.16, 0.22)$. The true DLT rates of the six scenarios are displayed in columns 3-7 of Table 3.3, with the operating characteristics of the three approaches summarized in the last three columns of Table 3.3. Instead of using a ten-point window as we did in Section 3.1, we focus our comparison of the three methods to the combinations with DLT rates that are exactly equal to $\Gamma = 0.30$ in scenarios 1-4 and $\Gamma = 0.40$ in scenarios 5-6.

From Table 3, we see that the gCRM does as well as (in scenarios 5 and 6) or better than (in scenarios 1-4) the methods of Yin and Yuan at identifying the MTC and always assigns a much higher percentage of patients to the MTC. We also see that the gCRM has comparable operating characteristics to the POCRM in scenarios 4 and 6 and better operating characteristics in scenarios 1, 2, and 5. However, the POCRM performs better than the gCRM in scenario 3. Overall, the results in Table Table 3.3 demonstrate that the gCRM has excellent operating characteristics and is a serious contender with existing methods. Our results also demonstrate that the gCRM works well regardless of the number of doses studied for each of the two agents. We also note that the gCRM uses a relatively small set of elicited information from which to design the entire study, while the POCRM requires a "skeleton" of elicited probabilities for DLT rates of all the dose combinations.

3.3.3 Assessing Sensitivity to Elicited DLT Probabilities

Recall that the results in Table 3.1 were based upon a first set of elicited values in which p^*_{1k} were equal to the first column of the true DLT rates in scenario A and p^*_{j1} were equal to the first row of the true DLT rates in scenario A. We also examined the operating characteristics of the gCRM in scenarios C, E, F, and G when using two additional sets of elicited DLT rates supplied by the investigator. The second set reverses the values of p^*_{1k} and p^*_{j1}, i.e. p^*_{1k} are equal to the first row of the true DLT rates in scenario A and p^*_{j1} are equal to the first column of the true DLT rates in scenario A. The third set has p^*_{1k} and p^*_{j1} equal to the first column and first row, respectively, of the true DLT rates in scenario E. All other hyperparameter values remain unchanged from those in Section 3.1. Operating characteristics resulting from 2,000 simulations are shown in Table 3.4, in which columns 3-6 are the values from Table 3.1 using the first set of elicited values, columns 7-10 correspond to the second set of elicited values, and columns 11-14 correspond to the third set of elicited values.

From Table 3.4, we see that the operating characteristics resulting from the three different sets of elicited values are fairly comparable among all the scenarios, although the results certainly reflect some sensitivity of the gCRM to the elicited values. For example, in scenario C, combination $(2, 1)$ is selected more often as the MTC and assigned to more patients when using the first set of elicited values and combination $(1, 2)$ is selected more often as the MTC and assigned to more patients when using the third set of elicited values. Furthermore, we see that the operating characteristics in scenario E are improved when using the third set of elicited values, which are based on the true DLT rates in scenario E, than the other two sets of elicited values.

3.4 Discussion

The gCRM is a logical extension of the original CRM for single-agent Phase I trials to the study of two agents, in which we extend the logistic model used in the traditional CRM to proportional odds logistic models. However, the gCRM only assumes proportional odds of DLT for adjacent doses of one of the agents (Agent A) when combined with the same dose of the other agent (Agent B). The odds of DLT among adjacent doses of Agent B, when combined with the same dose of Agent A, are not necessarily required to be proportional nor monotonic, although monotonicity is expected and can be promoted through appropriate selection of the prior variance parameter. In essence, we have reduced the problem of estimating $m \times n$ probabilities of DLT to the estimation of $n + 1$ model parameters. This fact motivates our suggestion that the label "Agent B" be given to the agent with fewer dose levels being studied, as this will minimize the number of parameters that need to estimated from the traditionally small sample sizes used in Phase I trials.

One of the obvious limitations in existing Phase I trial designs of two agents, including our work, is the implicit assumption that each patient must complete their follow-up for DLT before the next patient can be enrolled. Like the time-to-event CRM (TITE-CRM) (*Cheung and Chappell*, 2000), we are working on methods to incorporate results for patients with incomplete follow-up into the decision rule for assignment of combinations. One of the interesting issues is that both agents will most often not be given at the same time, thereby requiring methods that account for the differing amounts of follow-up that could occur for the two agents.

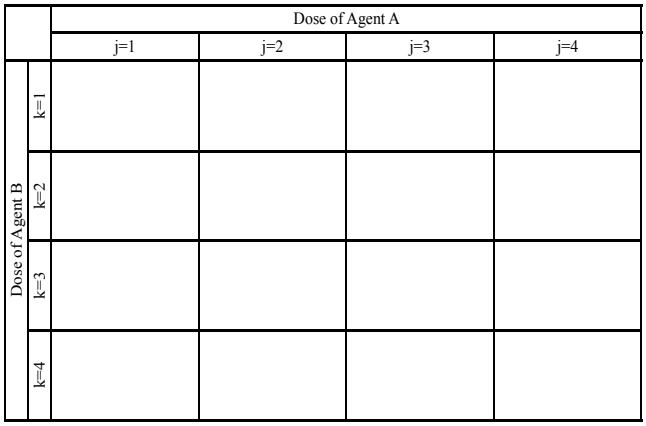

Figure 3.1: Schematic representation of Phase I trial of two agents, with four doses of each agent.

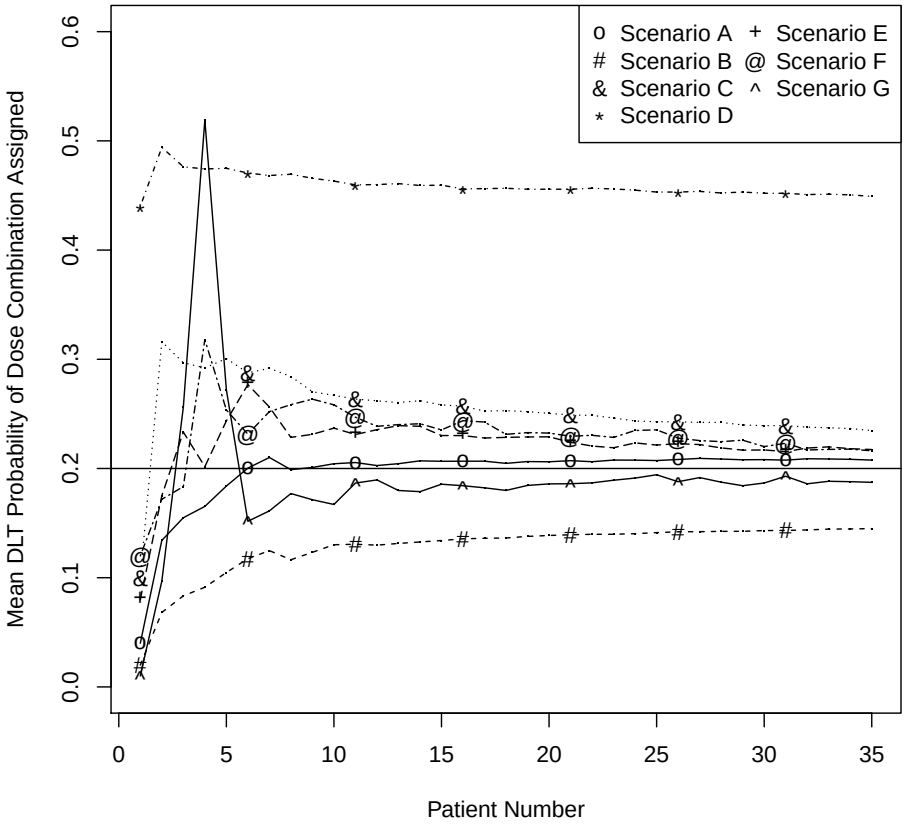

Figure 3.2: Average dose assigned to each patient in simulations.

Table 3.1: Operating characteristics of the gCRM in seven hypothetical scenarios. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate that is within 10 points of the desired DLT rate $\Gamma = 0.20$.

		True DLT Rates				Percentage of simulations selected as MTC				Mean percentage of patients assigned			
						Agent A							
Scenario	Agent B	1	2	3	4	1	2	3	4	1	2	3	4
A	1	4	8	**12**	**16**	0	2	**4**	**8**	6	4	**4**	**6**
	2	**10**	**14**	**18**	**22**	**1**	**8**	**11**	**10**	**3**	**10**	**7**	**7**
	3	**16**	**20**	**24**	**28**	**5**	**10**	**10**	**6**	**5**	**6**	**10**	**6**
	4	**22**	**26**	**30**	34	**5**	**7**	**5**	6	**4**	**4**	**5**	13
B	1	2	4	6	8	0	0	0	2	4	1	1	2
	2	5	7	9	**11**	0	1	2	**3**	1	5	2	**3**
	3	8	**10**	**12**	**14**	0	**2**	**5**	**8**	1	**2**	**8**	**6**
	4	**11**	**13**	**15**	**17**	**1**	**4**	**10**	**62**	**2**	**3**	**7**	**53**
C	1	**10**	**20**	**30**	40	**21**	**39**	**13**	3	**29**	**25**	**9**	4
	2	**25**	35	45	55	**13**	2	0	0	**12**	8	1	1
	3	40	50	60	70	1	0	0	0	3	0	3	0
	4	55	65	75	85	0	0	0	0	0	0	0	1
D	1	44	48	52	56	5	0	0	0	31	2	1	1
	2	50	54	58	62	0	0	0	0	1	3	0	0
	3	56	60	64	68	0	0	0	0	0	0	1	0
	4	62	66	70	74	0	0	0	0	0	0	0	0
E	1	8	**18**	**28**	38	7	**18**	**10**	3	14	**11**	**6**	4
	2	9	**19**	**29**	39	4	**10**	**8**	3	5	**10**	**6**	4
	3	**10**	**20**	**30**	40	**5**	**9**	**4**	2	**6**	**5**	**8**	3
	4	**11**	**21**	31	41	**5**	**5**	3	1	**4**	**3**	3	7
F	1	**12**	**13**	**14**	**15**	**8**	**9**	**8**	**15**	**16**	**9**	**6**	**10**
	2	**16**	**18**	**20**	**22**	**7**	**14**	**12**	**11**	**8**	**15**	**9**	**8**
	3	44	45	46	47	5	2	1	0	5	2	5	1
	4	50	52	54	55	1	0	0	0	2	0	0	2
G	1	1	2	3	4	0	0	0	2	4	1	1	4
	2	4	**10**	**15**	**20**	0	**9**	**18**	**24**	2	**12**	**12**	**14**
	3	6	**15**	**30**	45	3	**17**	**15**	3	6	**12**	**15**	4
	4	**10**	**30**	50	80	**5**	**3**	1	0	**7**	**3**	1	3

Table 3.2: Comparison of gCRM operating characteristics to those in Braun and Wang (2010) in three scenarios. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate within 10 points of the desired DLT rate $\Gamma = 0.20$.

		True DLT Rates				gCRM				Braun & Wang			
						Agent A							
Scenario	Agent B	1	2	3	4	1	2	3	4	1	2	3	4
		Percentage of simulations selected as MTC											
C	1	**10**	**20**	**30**	40	**21**	**39**	**13**	3	**15**	**27**	**9**	3
	2	**25**	35	45	55	**13**	2	0	0	**20**	3	0	0
	3	40	50	60	70	1	0	0	0	0	0	0	0
	4	55	65	75	85	0	0	0	0	0	0	0	0
E	1	8	**18**	**28**	38	7	**18**	**10**	3	2	**9**	**6**	5
	2	9	**19**	**29**	39	4	**10**	**8**	3	8	**11**	**11**	3
	3	**10**	**20**	**30**	40	**5**	**9**	**4**	2	**6**	**12**	**6**	2
	4	**11**	**21**	31	41	**5**	**5**	3	1	**5**	**4**	3	2
F	1	**12**	**13**	**14**	**15**	**8**	**9**	**8**	**15**	**6**	**8**	**9**	**7**
	2	**16**	**18**	**20**	**22**	**7**	**14**	**12**	**11**	**16**	**14**	**6**	**1**
	3	44	45	46	47	5	2	1	0	11	4	1	0
	4	50	52	54	55	1	0	0	0	2	0	0	0
		Mean percentage of subjects assigned											
C	1	**10**	**20**	**30**	40	**29**	**25**	**9**	4	**23**	**21**	**8**	3
	2	**25**	35	45	55	**12**	8	1	1	**14**	9	2	1
	3	40	50	60	70	3	0	3	0	6	2	3	0
	4	55	65	75	85	0	0	0	1	1	0	0	1
E	1	8	**18**	**28**	38	14	**11**	**6**	4	10	**8**	**5**	3
	2	9	**19**	**29**	39	5	**10**	**6**	4	6	**12**	**7**	3
	3	**10**	**20**	**30**	40	**6**	**5**	**8**	3	**5**	**8**	**9**	3
	4	**11**	**21**	31	41	**4**	**3**	3	7	**4**	**3**	3	8
F	1	**12**	**13**	**14**	**15**	**16**	**9**	**6**	**10**	**14**	**9**	**7**	**5**
	2	**16**	**18**	**20**	**22**	**8**	**15**	**9**	**8**	**11**	**14**	**7**	**2**
	3	44	45	46	47	5	2	5	1	8	5	5	1
	4	50	52	54	55	2	0	0	2	2	1	1	3

Table 3.3: Comparison of the operating characteristics of the gCRM to the designs of Yin and Yuan (2009a; 2009b) and the POCRM of Wages, Conaway, O'Quigley (2011a). All values are percentages multiplied by 100. Boldfaced numbers correspond to dose combinations with a DLT rate equal to the desired DLT rate.

		Agent A						Correct	Observed	Patients
Scenario	Agent B	1	2	3	4	5	Method	Recommend	DLTs	at MTC
1	1	6	8	10	15	•	Yin & Yuan	41	28	13
	2	10	12	**30**	45	•	POCRM	53	31	39
	3	15	**30**	50	60	•	gCRM	66	31	44
	4	50	55	60	70	•				
2	1	8	12	16	18	•	Yin & Yuan	53	28	19
	2	10	15	**30**	45	•	POCRM	50	31	39
	3	12	**30**	50	55	•	gCRM	65	30	46
	4	**30**	50	55	60	•				
3	1	6	8	10	**30**	50	Yin & Yuan	63	28	21
	2	12	16	**30**	50	55	POCRM	86	31	69
	3	15	**30**	50	55	60	gCRM	74	31	49
4	1	6	10	15	**30**	50	Yin & Yuan	56	29	19
	2	10	**30**	50	70	80	POCRM	68	33	43
	3	50	60	70	80	90	gCRM	71	31	48
5	1	24	**40**	47	56	64	Yin & Yuan	44	37	23
	2	**40**	45	59	67	74	POCRM	38	46	27
	3	48	59	68	75	81	gCRM	46	43	37
	4	54	67	75	81	86				
6	1	18	29	**40**	47	56	Yin & Yuan	48	35	22
	2	27	**40**	45	59	67	POCRM	45	44	33
	3	**40**	49	59	68	75	gCRM	46	42	35
	4	49	58	68	75	81				

Table 3.4: Operating characteristics of the gCRM using three different sets of elicited DLT probabilities. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate within 10 points of the desired DLT rate $\Gamma = 0.20$.

		First set of elicited values				Second set of elicited values				Third set of elicited values			
		Agent A											
Scenario	Agent B	1	2	3	4	1	2	3	4	1	2	3	4
		Percentage of simulations selected as MTC											
C	1	**21**	**39**	**13**	3	**20**	**32**	**8**	2	**32**	**13**	**3**	0
	2	**13**	2	0	0	**24**	2	0	0	**34**	2	0	0
	3	1	0	0	0	2	0	0	0	5	0	0	0
	4	0	0	0	0	0	0	0	0	1	0	0	0
E	1	7	**18**	**18**	3	4	**16**	**11**	4	3	**11**	**7**	2
	2	4	**10**	**8**	3	6	**11**	**5**	3	5	**15**	**6**	2
	3	**5**	**9**	**4**	2	**6**	**9**	**4**	2	**6**	**14**	**5**	2
	4	**5**	**5**	3	1	**7**	**5**	3	1	**7**	**9**	3	1
F	1	**8**	**9**	**8**	**15**	**10**	**6**	**7**	**11**	**14**	**9**	**7**	**6**
	2	**7**	**14**	**12**	**11**	**11**	**13**	**13**	**10**	**21**	**9**	**9**	**6**
	3	5	2	1	0	6	2	1	0	6	3	1	0
	4	1	0	0	0	1	0	0	0	2	0	0	0
G	1	0	0	0	2	0	0	0	4	0	1	3	8
	2	0	**9**	**18**	**24**	0	**7**	**21**	**22**	0	**5**	**16**	**16**
	3	3	**17**	**15**	3	1	**20**	**15**	2	1	**20**	**15**	2
	4	**5**	**3**	1	0	**4**	**2**	0	0	**4**	**7**	2	0
		Mean percentage of patients assigned											
C	1	**29**	**25**	**9**	4	**27**	**21**	**9**	3	**36**	**7**	**2**	0
	2	**12**	8	1	1	**19**	7	1	1	**25**	7	1	0
	3	3	0	3	0	3	0	2	0	8	1	3	0
	4	0	0	0	1	1	0	0	1	3	0	0	1
E	1	14	**11**	**6**	4	11	**11**	**9**	4	11	**7**	**5**	2
	2	5	**10**	**6**	4	8	**10**	**5**	3	6	**13**	**5**	3
	3	**6**	**5**	**8**	3	**4**	**6**	**7**	3	**6**	**9**	**7**	3
	4	**4**	**3**	3	7	**4**	**3**	3	7	**7**	**6**	3	7
F	1	**16**	**9**	**6**	**10**	**17**	**6**	**7**	**8**	**21**	**6**	**5**	**4**
	2	**8**	**15**	**9**	**8**	**9**	**12**	**11**	**7**	**16**	**10**	**6**	**5**
	3	5	2	5	1	4	3	5	1	7	3	5	1
	4	2	0	0	2	2	1	1	3	3	1	1	3
G	1	4	1	1	4	4	1	3	6	4	3	5	6
	2	2	**12**	**12**	**14**	1	**11**	**16**	**14**	1	**10**	**11**	**11**
	3	6	**12**	**15**	4	2	**12**	**16**	3	3	**11**	**14**	4
	4	**7**	**3**	1	3	**4**	**3**	2	3	**5**	**6**	2	3

Table 3.5: Operating characteristics of the gCRM when assuming the priors of the hyperparameters are uniformly distributed in their respective recommended range. All values are multiplied by 100. Boldfaced numbers correspond to dose combinations with a true DLT rate that is within 10 points of the desired DLT rate $\Gamma = 0.20$. Simulation size is 1000, draw size is 5000 with a burn-in of 1000.

		True DLT Rates				Percentage of simulations selected as MTC				Mean percentage of patients assigned			
						Agent A							
Scenario	Agent B	1	2	3	4	1	2	3	4	1	2	3	4
A	1	4	8	**12**	**16**	0	3	**4**	**7**	6	4	**3**	**5**
	2	**10**	**14**	**18**	**22**	**2**	**8**	**12**	**20**	**4**	**10**	**8**	**7**
	3	**16**	**20**	**24**	**28**	**5**	**11**	**11**	**6**	**5**	**6**	**11**	**6**
	4	**22**	**26**	**30**	34	**4**	**6**	**4**	5	**4**	**4**	**5**	13
B	1	2	4	6	8	0	0	0	0	4	1	0	1
	2	5	7	9	**11**	0	0	2	**2**	1	5	2	**2**
	3	8	**10**	**12**	**14**	1	**2**	**5**	**7**	2	**2**	**7**	**6**
	4	**11**	**13**	**15**	**17**	**1**	**4**	**10**	**64**	**2**	**3**	**8**	**54**
C	1	**10**	**20**	**30**	40	**20**	**37**	**12**	2	**26**	**26**	**8**	4
	2	**25**	35	45	55	**14**	2	0	0	**14**	8	1	1
	3	40	50	60	70	2	0	0	0	3	0	2	0
	4	55	65	75	85	0	0	0	0	1	0	0	1
D	1	44	48	52	56	6	1	0	0	31	3	1	0
	2	50	54	58	62	0	0	0	0	1	2	0	0
	3	56	60	64	68	0	0	0	0	0	0	1	0
	4	62	66	70	74	0	0	0	0	0	0	0	0
E	1	8	**18**	**28**	38	6	**14**	**9**	4	13	**9**	**6**	3
	2	9	**19**	**29**	39	5	**10**	**10**	4	6	**11**	**6**	3
	3	**10**	**20**	**30**	40	**6**	**8**	**4**	2	**7**	**5**	**7**	3
	4	**11**	**21**	31	41	**6**	**4**	3	2	**5**	**3**	3	7
F	1	**12**	**13**	**14**	**15**	**7**	**9**	**7**	**14**	**15**	**9**	**5**	**9**
	2	**16**	**18**	**20**	**22**	**10**	**16**	**12**	**10**	**11**	**16**	**8**	**7**
	3	44	45	46	47	5	2	1	0	6	2	5	1
	4	50	52	54	55	1	0	0	0	2	1	1	2
G	1	1	2	3	4	0	0	0	1	4	1	1	4
	2	4	**10**	**15**	**20**	0	**9**	**20**	**19**	2	**12**	**13**	**13**
	3	6	**15**	**30**	45	3	**20**	**16**	3	5	**11**	**16**	4
	4	**10**	**30**	50	80	**5**	**3**	1	0	**7**	**3**	2	3

Table 3.6: Simulation results of the gCRM using the power model. True DLT probabilities, and simulation results for scenarios A through F. All values are multiplied by 100. Bold faced numbers correspond to dose combinations with a true DLT rate exactly equal to the target DLT rate.

		Agent A											
		True Dlt Rates				Percentage of simulations selected as MTC				Mean percentage of patients assigned			
Scenario	Agent B	1	2	3	4	1	2	3	4	1	2	3	4
A	1	4	8	**12**	**16**	0	0	**1**	**13**	3	1	**2**	**10**
	2	**10**	**14**	**18**	**22**	**2**	**8**	**10**	**15**	**4**	**9**	**8**	**13**
	3	**16**	**20**	**24**	**28**	**3**	**11**	**5**	**9**	**3**	**4**	**8**	**12**
	4	**22**	**26**	**30**	34	**5**	**5**	**8**	3	**3**	**3**	**6**	11
B	1	2	4	6	8	0	0	1	4	3	0	0	5
	2	5	7	9	**11**	0	0	2	**13**	0	4	2	**11**
	3	8	**10**	**12**	**14**	0	**2**	**5**	**24**	0	**2**	**7**	**18**
	4	**11**	**13**	**15**	**17**	**0**	**5**	**7**	**37**	**1**	**3**	**7**	**35**
C	1	**10**	**20**	**30**	40	**10**	**24**	**16**	4	**11**	**16**	**14**	12
	2	**25**	35	45	55	**16**	4	1	0	**13**	8	3	2
	3	40	50	60	70	0	0	0	0	1	0	2	0
	4	55	65	75	85	0	0	0	0	0	0	0	1
D	1	44	48	52	56	3	0	0	0	17	5	3	2
	2	50	54	58	62	0	0	0	0	3	2	1	1
	3	56	60	64	68	0	0	0	0	0	0	1	0
	4	62	66	70	74	0	0	0	0	0	0	0	0
E	1	8	**18**	**28**	38	1	**9**	**7**	0	4	**4**	**5**	4
	2	9	**19**	**29**	39	8	**20**	**13**	3	8	**17**	**12**	8
	3	**10**	**20**	**30**	40	**5**	**10**	**3**	3	**4**	**6**	**7**	6
	4	**11**	**21**	31	41	**4**	**6**	3	0	**1**	**3**	4	5
F	1	**12**	**13**	**14**	**15**	**0**	**2**	**5**	**19**	**5**	**2**	**3**	**15**
	2	**16**	**18**	**20**	**22**	**9**	**21**	**9**	**22**	**9**	**15**	**11**	**18**
	3	44	45	46	47	5	0	0	0	4	2	5	3
	4	50	52	54	55	0	0	0	0	1	1	1	2
G	1	1	2	3	4	0	0	0	4	4	1	1	4
	2	4	**10**	**15**	**20**	1	**8**	**12**	**25**	3	**4**	**7**	**19**
	3	6	**15**	**30**	45	3	**19**	**21**	1	4	**17**	**19**	5
	4	**10**	**30**	50	80	**2**	**4**	0	0	**1**	**2**	2	3

Table 3.7: True DLT probabilities, and simulation results for scenarios with four levels of agent A and three levels of agent B. All values are multiplied by 100. Scenario A1 boundary DLT rates are used in all seven scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$.

		Agent A											
		True Dlt Rates				Percentage of simulations selected as MTC				Mean percentage of patients assigned			
Scenario	Agent B	1	2	3	4	1	2	3	4	1	2	3	4
A1	1	4	8	**12**	**16**	0	2	**6**	**11**	6	5	**4**	**7**
	2	**10**	**14**	**18**	**22**	**2**	**7**	**13**	**12**	**3**	**9**	**9**	**8**
	3	**16**	**20**	**24**	**28**	**4**	**13**	**14**	**17**	**4**	**7**	**12**	**14**
B1	1	2	4	6	8	0	0	0	1	4	1	1	2
	2	5	7	9	**11**	1	5	4	**5**	1	5	4	**5**
	3	8	**10**	**12**	**14**	2	**4**	**13**	**71**	2	**4**	**13**	**59**
C1	1	**10**	**20**	**30**	40	**20**	**40**	**13**	3	**28**	**26**	**10**	5
	2	**25**	35	45	55	**12**	2	0	0	**11**	8	2	1
	3	40	50	60	70	1	0	0	0	3	1	2	1
D1	1	44	48	52	56	5	0	0	0	31	2	1	0
	2	50	54	58	62	0	0	0	0	1	2	0	0
	3	56	60	64	68	0	0	0	0	0	0	1	0
E1	1	8	**18**	**28**	38	6	**24**	**13**	4	13	**15**	**8**	5
	2	9	**19**	**29**	39	5	**12**	**9**	4	7	**11**	**8**	4
	3	**10**	**20**	**30**	40	**5**	**9**	**5**	1	**6**	**6**	**8**	9
F1	1	**12**	**13**	**14**	**15**	**9**	**8**	**7**	**15**	**16**	**8**	**6**	**10**
	2	**16**	**18**	**20**	**22**	**7**	**16**	**13**	**14**	**7**	**14**	**10**	**9**
	3	44	45	46	47	5	2	1	0	5	3	5	4
G1	1	1	2	3	4	0	0	0	2	4	1	1	5
	2	4	**10**	**15**	**20**	0	**9**	**20**	**22**	2	**11**	**16**	**15**
	3	6	**15**	**30**	45	3	**24**	**17**	3	5	**12**	**16**	11

Table 3.8: True DLT probabilities, and simulation results for scenarios with five levels of agent A and three levels of agent B. All values are multiplied by 100. Scenario A2 boundary DLT rates are used in all seven scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$.

		Agent A														
		True Dlt Rates					Percentage of simulations selected as MTC					Mean percentage of patients assigned				
Scenario	Agent B	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
A2	1	4	8	**12**	**16**	**20**	0	2	**6**	**8**	**10**	6	4	**5**	**5**	**8**
	2	**10**	**14**	**18**	**22**	**26**	**2**	**5**	**12**	**10**	**7**	**3**	**8**	**7**	**8**	**7**
	3	**16**	**20**	**24**	**28**	32	**2**	**7**	**11**	**9**	8	**3**	**4**	**8**	**9**	15
B2	1	2	4	6	8	**11**	0	0	0	1	**3**	4	1	1	1	**3**
	2	5	7	9	**11**	**13**	0	0	2	**3**	**6**	1	4	2	**4**	**5**
	3	8	**10**	**12**	**14**	**17**	0	**1**	**5**	**10**	**69**	1	**2**	**6**	**11**	**53**
C2	1	**10**	**20**	**30**	40	50	**21**	**40**	**16**	3	0	**29**	**27**	**11**	4	2
	2	**25**	35	45	55	65	**10**	1	0	0	0	**10**	7	1	1	0
	3	40	50	60	70	80	1	0	0	0	0	2	1	2	1	0
D2	1	44	48	52	56	60	6	0	0	0	0	31	2	1	0	0
	2	50	54	58	62	66	0	0	0	0	0	1	3	0	0	0
	3	56	60	64	68	72	0	0	0	0	0	0	0	1	0	0
E2	1	8	**18**	**28**	38	48	6	**26**	**18**	5	1	13	**16**	**11**	6	3
	2	9	**19**	**29**	39	49	6	**10**	**7**	2	1	6	**10**	**5**	4	2
	3	**10**	**20**	**30**	40	50	**6**	**5**	**2**	1	0	**6**	**4**	**4**	4	3
F2	1	**12**	**13**	**14**	**15**	**16**	**8**	**9**	**9**	**9**	**15**	**16**	**8**	**6**	**6**	**10**
	2	**16**	**18**	**20**	**22**	**24**	**5**	**8**	**10**	**8**	**8**	**7**	**12**	**7**	**7**	**6**
	3	44	45	46	47	48	4	2	1	0	0	4	2	4	3	2
G2	1	1	2	3	4	5	0	0	0	1	6	4	1	1	2	10
	2	4	**10**	**15**	**20**	**30**	0	**6**	**23**	**29**	**14**	1	**9**	**13**	**19**	**12**
	3	6	**15**	**30**	45	70	2	**9**	**7**	2	0	4	**7**	**8**	6	3

Table 3.9: True DLT probabilities, and simulation results for scenarios with five levels of agent A and four levels of agent B. All values are multiplied by 100. Scenario A3 boundary DLT rates are used in all seven scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$. A sample size of 35 patients is used in all seven scenarios.

		Agent A														
		True Dlt Rates					Percentage of simulations selected as MTC					Mean percentage of patients assigned				
Scenario	Agent B	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
A3	1	4	8	**12**	**16**	**20**	0	2	**6**	**8**	**11**	6	4	**4**	**6**	**8**
	2	**10**	**14**	**18**	**22**	**26**	**2**	**6**	**9**	**8**	**6**	**3**	**8**	**5**	**7**	**5**
	3	**16**	**20**	**24**	**28**	32	**3**	**6**	**8**	**6**	3	**4**	**4**	**7**	**6**	4
	4	**22**	**26**	**30**	34	38	**3**	**4**	**3**	3	2	**2**	**2**	**2**	5	7
B3	1	2	4	6	8	**11**	0	0	0	1	**3**	4	1	1	1	**3**
	2	5	7	9	**11**	**13**	0	0	2	**3**	**5**	1	4	1	**3**	**4**
	3	8	**10**	**12**	**14**	**17**	0	**1**	**3**	**6**	**7**	1	**1**	**5**	**5**	**6**
	4	**11**	**13**	**15**	**17**	**19**	**1**	**3**	**5**	**11**	**49**	**1**	**1**	**3**	**10**	**42**
C3	1	**10**	**20**	**30**	40	50	**19**	**40**	**15**	3	0	**28**	**26**	**10**	4	2
	2	**25**	35	45	55	65	**11**	2	0	0	0	**11**	7	1	1	0
	3	40	50	60	70	80	1	0	0	0	0	3	0	3	0	0
	4	55	65	75	85	95	0	0	0	0	0	0	0	0	1	0
D3	1	44	48	52	56	60	6	0	0	0	0	31	2	1	0	0
	2	50	54	58	62	66	0	0	0	0	0	1	2	0	0	0
	3	56	60	64	68	72	0	0	0	0	0	0	0	1	0	0
	4	62	66	70	74	78	0	0	0	0	0	0	0	0	0	0
E3	1	8	**18**	**28**	38	48	7	**22**	**16**	5	1	14	**13**	**10**	6	3
	2	9	**19**	**29**	39	49	4	**9**	**5**	2	1	5	**10**	**3**	4	2
	3	**10**	**20**	**30**	40	50	**5**	**5**	**3**	1	0	**5**	**3**	**5**	3	1
	4	**11**	**21**	31	41	51	**4**	**2**	1	0	0	**3**	**1**	1	2	3
F3	1	**12**	**13**	**14**	**15**	**16**	**8**	**7**	**9**	**10**	**14**	**16**	**8**	**6**	**6**	**10**
	2	**16**	**18**	**20**	**22**	**24**	**6**	**9**	**10**	**8**	**7**	**7**	**12**	**6**	**7**	**5**
	3	44	45	46	47	48	4	2	1	0	0	4	1	4	2	1
	4	50	52	54	56	58	0	0	0	0	0	1	0	0	1	1
G3	1	1	2	3	4	5	0	0	0	1	6	4	1	1	2	9
	2	4	**10**	**15**	**20**	**30**	0	**6**	**21**	**26**	**15**	2	**9**	**10**	**17**	**10**
	3	6	**15**	**30**	45	70	2	**10**	**7**	2	0	5	**7**	**9**	4	1
	4	**10**	**30**	50	80	99	**4**	**2**	0	0	0	**5**	**2**	1	2	0

Table 3.10: Simulations results of different combinations of μ, θ and σ^2 using the proportional odds logistic model of the gCRM. All values are multiplied by 100. Scenario A DLT rates are assumed as the true DLT rates in all the 11 scenarios. Bold faced numbers correspond to dose combinations with a true DLT rate that is no more than 10 points away from the target toxicity rate $\Gamma = 0.2$. Simulation size is 2000, draw size is 5000 with a burn-in of 2000. A sample size of 35 patients is used in all seven scenarios.

		Agent A									Agent A							
Agent B	Hyper-params	Percentage of Simulations Selected as MTC				Mean percentage of Patients assigned				Hyper-Params	Percentage of Simulations Selected as MTC				Mean Percentage of Patients assigned			
		1	2	3	4	1	2	3	4		1	2	3	4	1	2	3	4
1	$\mu = -8$	0	2	**4**	**7**	7	4	**3**	**5**	$\mu = -8.5$	0	3	**5**	**9**	6	4	**4**	**7**
2	$\beta = 5$	**4**	**8**	**11**	**8**	**7**	**8**	**8**	**5**	$\beta = 4.5$	**2**	**9**	**10**	**9**	**3**	**10**	**7**	**6**
3	$\sigma^2 = 0.3$	**7**	**10**	**10**	**7**	**5**	**9**	**8**	**4**	$\sigma^2 = 0.5$	**5**	**9**	**9**	**7**	**5**	**6**	**11**	**6**
4		**4**	**7**	**7**	5	**3**	**5**	**1**	11		**4**	**7**	**6**	5	**3**	**4**	**5**	13
1	$\mu = -8$	0	2	**5**	**8**	7	4	**3**	**6**	$\mu = -8.5$	0	3	**5**	**9**	6	4	**4**	**6**
2	$\beta = 5$	**3**	**8**	**11**	**8**	**4**	**9**	**7**	**6**	$\beta = 5.5$	**1**	**7**	**11**	**9**	**3**	**9**	**7**	**7**
3	$\sigma^2 = 0.4$	**5**	**10**	**10**	**7**	**5**	**7**	**10**	**6**	$\sigma^2 = 0.5$	**4**	**10**	**10**	**8**	**4**	**6**	**11**	**7**
4		**4**	**7**	**7**	4	**3**	**4**	**6**	12		**5**	**8**	**4**	5	**4**	**4**	**5**	13
1	$\mu = -8$	0	3	**5**	**11**	6	4	**4**	**8**	$\mu = -8.5$	0	2	**5**	**9**	6	4	**4**	**6**
2	$\beta = 5$	**2**	**7**	**11**	**11**	**2**	**9**	**7**	**8**	$\beta = 5$	**2**	**8**	**12**	**9**	**3**	**10**	**7**	**6**
3	$\sigma^2 = 0.6$	**4**	**9**	**9**	**7**	**4**	**5**	**10**	**7**	$\sigma^2 = 0.5$	**4**	**10**	**8**	**6**	**3**	**6**	**10**	**6**
4		**6**	**6**	**5**	4	**3**	**4**	**5**	12		**4**	**7**	**6**	5	**4**	**4**	**5**	14
1	$\mu = -7.5$	0	2	**5**	**9**	6	4	**4**	**6**	$\mu = -8$	0	3	**5**	**9**	6	4	**4**	**6**
2	$\beta = 4.5$	**2**	**8**	**11**	**10**	**4**	**9**	**7**	**7**	$\beta = 4.5$	**2**	**8**	**11**	**10**	**3**	**10**	**7**	**7**
3	$\sigma^2 = 0.5$	**5**	**10**	**8**	**7**	**5**	**7**	**9**	**6**	$\sigma^2 = 0.5$	**4**	**10**	**10**	**6**	**5**	**6**	**10**	**6**
4		**5**	**6**	**5**	5	**3**	**4**	**6**	13		**4**	**8**	**5**	5	**4**	**4**	**5**	13
1	$\mu = -7.5$	0	3	**4**	**9**	6	4	**4**	**6**	$\mu = -8$	0	3	**5**	**8**	6	4	**4**	**6**
2	$\beta = 5.5$	**2**	**8**	**12**	**10**	**4**	**9**	**8**	**7**	$\beta = 5.5$	**1**	**9**	**11**	**10**	**3**	**10**	**8**	**7**
3	$\sigma^2 = 0.5$	**4**	**10**	**9**	**7**	**5**	**7**	**9**	**6**	$\sigma^2 = 0.5$	**4**	**10**	**10**	**6**	**5**	**6**	**11**	**6**
4		**4**	**7**	**6**	5	**3**	**4**	**6**	12		**5**	**6**	**5**	4	**4**	**4**	**5**	12
1	$\mu = -7.5$	0	2	**4**	**9**	6	4	**3**	**6**									
2	$\beta = 5$	**3**	**9**	**11**	**11**	**4**	**9**	**8**	**7**									
3	$\sigma^2 = 0.5$	**5**	**9**	**8**	**7**	**5**	**7**	**9**	**6**									
4		**5**	**7**	**6**	4	**3**	**4**	**6**	12									

CHAPTER IV

A Time-To-Event Approach to the gCRM

4.1 Introduction

Similar to most of the existing methods in Phase I oncology trials, the gCRM as proposed in Chapter 3 is built on the assumption that each patient is completely followed before the next available patient can be assigned to a dose combination. This is because the outcome of interest, although truly assessed over a follow-up interval $[0, T]$, is treated as binary, indicating whether or not a patient experienced a DLT in that interval. A patient is considered to be fully followed when either they experience a DLT or reach time T without experiencing a DLT. Thus, any patient with a total follow-up of $0 < t < T$ who has not experienced a DLT has yet to have an observed outcome and would contribute nothing to a likelihood of binary outcomes. Traditional methods allow for two solutions to this dilemma when a new patient might be enrolled.

First, we could simply ignore any patients with incomplete follow-up and base dose assignment decisions solely using data from completely followed patients. This was the suggestion made by the original creators of the CRM (*O'Quigley et al.*, 1990) , which is statistically unappealing since it is inefficient and does not incorporate all

the information gathered in the study. Second, we could restrict enrollment of a new patient to occur only when all previously enrolled patients have completed follow-up. However, such a restriction is unrealistic in practice because most patients eligible for Phase I trials are gravely ill and require treatment as soon as they are eligible to receive it.

While little research has been done on incorporation of partial follow-up in two agent Phase I trial designs, there are two existing methods for one-agent Phase I trials: the time-to-event CRM (TITE-CRM) by *Cheung and Chappell* (2000), and the EM-CRM by *Yuan and Yin* (2011b). Recall that we have a single-parameter dose-toxicity model $F(\beta; d)$ that computes a probability of DLT by time T for a patient receiving dose d. In a sample of N patients, we let $D_{[i]}$ denote the dose assigned to patient $i, i = 1, 2, \ldots N$ and Y_i be the indicator of DLT by time T for patient i. If we let $\mathcal{D}$ denote the data from all N patients, then the likelihood used for estimating β would be

$$L_N(\beta \mid \mathcal{D}) = \prod_i F(\beta; D_{[i]})^{Y_i}[1 - F(\beta; D_{[i]})]^{(1-Y_i)}. \tag{4.1}$$

The TITE-CRM generalizes the traditional CRM by computing a weight function $0 \leq w(t) \leq 1$ for each partially-followed patient that encompasses the probability that patients who currently have not had a DLT will fail to experience DLT by the end of their follow-up. Thus, one needs a model for the distribution of times to DLT to compute these weights. Given the difficulty in determining this model and any additional parameters that would need to be estimated, a standard implementation of the TITE-CRM uses a weight for each partially-followed patient that is their proportion of completed follow-up, which assumes a uniform distribution of DLT times in the interval $[0, T]$ that is the same for each dose. As a result, β would be estimated us-

ing the weighted likelihood $L_N(\beta \mid \mathcal{D}) \propto \prod_i F(\beta; D_{[i]})^{Y_i}[1 - w_i F(\beta; D_{[i]})]^{(1-Y_i)}$, where $w_i = t_i/T$ for a patient with follow-up t_i. *Cheung and Chappell* (2000) also propose an alternate weight function if late-onset DLTs are expected and more generalized approach to accommodating late-onset DLTs was proposed by *Braun* (2006).

The EM-CRM is an extension to the maximum-likelihood version (as opposed to the Bayesian methods that we use) of the CRM (*Shen and O'Quigley*, 1996) that treats the unobserved outcomes of the partially-followed patients as missing data and uses the expectation-maximization (EM) algorithm (*Dempster et al.*, 1977) to compute a maximum-likelihood estimate of β. The authors let $s_1, s_2, ..., s_K$ denote the K unique times at which DLT has been observed in the study and also define $s_{K+1} = T$ for mathematical necessity. A discrete hazard function $\lambda_k = P(t = s_k \mid t >= s_k)$ is introduced for each s_k and is used to impute a probability of future DLT for partially-followed patients. Essentially, the EM-CRM computes a nonparametric estimate of the probability of DLT by time T given no DLT has yet been observed. For patient i with partial follow-up t_i, we can use this nonparametric estimate to impute a value for the missing Y_i as its expected value

$$\widetilde{Y}_i = \frac{F(\beta; D_{[i]}) \prod_{k:s_k<t_i}(1 - \lambda_k)}{1 - F(\beta; D_{[i]}) + F(\beta; D_{[i]}) \prod_{k:s_k<t_i}(1 - \lambda_k)},$$

for a given value of β. We then maximize the likelihood in Equation (4.1) with $\widetilde{Y}_i$ used in place of Y_i for each partially-followed patient to get an updated value of β. We then iteratively update the values of $\widetilde{Y}_i$ and β until convergence. Although a Bayesian form of the EM-CRM has never been formally proposed, such an approach is easily created from the Bayesian data augmentation ideas of *Tanner and Wong* (1987).

As the TITE-CRM generalizes the CRM, we propose a time-to-event version of the gCRM, which we call the weighted gCRM (wgCRM). As compared to the gCRM with partial follow-up information completely ignored, we expect the wgCRM will assign more patients to the MTC and identify the MTC more often at the end of the study. We discuss possible choices of the weight function as well as the implementation of the wgCRM in Section 4.2, present the simulation settings and results in Section 4.3, and present concluding remarks in Section 4.4.

4.2 Methods

4.2.1 Description of the gCRM

In Chapter 3, we have generalized the CRM, denoted as gCRM, to a Phase I trial of two agents in which there are m doses, $a_1 < a_2 < \ldots < a_m$, of Agent A, and n doses, $b_1 < b_2 < \ldots < b_n$, of Agent B. We let (j, k) represent the combination of Agent A at dose $a_j, j = 1, 2, \ldots, m$ and Agent B at dose $b_k, k = 1, 2, \ldots, n$. We let p_{jk} denote the true DLT probability for (j, k) and define the MTC, (j^*, k^*), as the combination with a DLT rate closest to the target rate Γ. For each value of k (each dose of Agent B), we adopt a traditional logistic regression model $\text{logit}(p_{jk}) = \alpha_k + \beta a_j$ describing how the probability of DLT varies with each dose of Agent A combined with the given dose of Agent B. Note that the intercept is now a parameter specific to the dose of Agent B under study, rather than a fixed value traditionally used in the CRM. In more general applications, this model is known as a proportional odds logistic regression model. We attempt to "join" the row-wise CRM models through a joint prior distribution that forces correlation among their intercept parameters. We note that like the traditional CRM, the dose values $a_1, a_2, \ldots, a_m$ of Agent A are not

actual clinical dose values, but are rescaled dose values selected to improve model fit. However, there is no need to compute rescaled dose values for Agent B because the doses of Agent B do not contribute directly to our model.

We assume $\log(\beta)$ has a normal distribution with mean ν and variance σ^2, denoted $\log(\beta) \sim \mathcal{N}(\nu, \sigma^2)$, and the intercept for the model at the lowest dose of Agent B, $\alpha_1 \sim \mathcal{N}(\mu, \sigma^2)$. Instead of directly modeling the joint distribution of the intercepts, we focus on the difference between intercepts corresponding to adjacent doses of Agent B. Specifically, we assume $\Delta_k = \alpha_k - \alpha_{k-1} \sim \mathcal{N}(\delta_k, \sigma^2), k = 2, 3, \ldots, n$. We note that the use of a normal distribution for the differences does allow for the possibility of non-increasing DLT rates among doses of Agent B for a fixed dose of Agent A. However, such a possibility is minimized if the value of σ^2 is kept sufficiently small so as to force a majority of the prior distribution above zero. If we let (j_i, k_i) denote the combination assigned to patient $i = 1, 2, \ldots M$ and Y_i denote whether or not a patient i has experienced DLT, then we have collective data $\mathcal{Y}_M = \{Y_1, Y_2, \ldots Y_M\}$ and $\mathcal{C}_M = \{(j_1, k_1), (j_2, k_2), \ldots, (j_M, k_M)\}$. This leads to a likelihood

$$L(\beta, \boldsymbol{\alpha} \mid \mathcal{Y}_M, \mathcal{C}_M) \propto \prod_{i=1}^{M} p_{j_i,k_i}^{Y_i} (1 - p_{j_i,k_i})^{1-Y_i} \tag{4.2}$$

that can be used to estimate values for β and $\boldsymbol{\alpha}$.

4.2.2 Incorporation of Partial Follow-up

The gCRM as just described earlier did not explicitly state that each patient is followed for DLT to time T after being enrolled. Furthermore, it was assumed that the inter-arrival times of patients were greater than T, so that we fully observe the DLT outcome of every enrolled patient. However, as we have stated in the Introduction

of this chapter, such an assumption is not likely to hold in most applications, necessitating methods that properly utilize information from partially followed patients. Our specific approach is as follows.

Out of a maximum planned sample size of N patients, suppose M patients, $1 \leq M \leq (N-1)$, have been enrolled in the study. Among these M patients, some patients are still under follow-up without experiencing DLT and the remaining patients are fully-followed, i.e. have either experienced DLT or been followed to time T without DLT. At the arrival of patient $(M+1)$, we will assign an individual weight w to each partially-followed subject that is a function of the length, $0 < t_i < T$ of their follow-up. Specific forms for w will be explained shortly. If we let $Y_i = 1$ for all patients experiencing DLT and $Y_i = 0$ for all other patients regardless of their length of follow-up, the likelihood for β and $\boldsymbol{\alpha}$ based upon the data of the first M patients is now

$$L(\beta, \boldsymbol{\alpha} \mid \mathcal{Y}_M, \mathcal{C}_M) \propto \prod_{i=1}^{M} p_{j_i,k_i}^{Y_i} [1 - w(t_i) p_{j_i,k_i}]^{1-Y_i},$$

which is a weighted form of the likelihood in Equation (4.2)

In the TITE-CRM, the weight given to each partially-followed patient was a function of their follow-up time $0 \leq t_i \leq T$. Any non-decreasing function $w(t_i)$ could be used as long as $w(0) = 0$ and $w(T) = 1$. Note that this latter constraint implicitly places complete weight, as desired, on subjects without DLT who have completed their follow-up. We also note that our weighted likelihood implicitly places full weight on subjects who experience DLT. As we stated earlier, the appropriate form of $w(t_i)$ would reflect the underlying distribution of the time to DLT, but for simplicity, the function $w(t_i) = t_i/T$ is conventionally used with the TITE-CRM, which assumes that DLT times are uniformly distributed over the interval $[0, T]$. However, in a trial

of two agents, say Agent A and Agent B, there is no reason to assume that both agents will be given simultaneously, as in many settings, one agent, say Agent A, would be given immediately upon enrollment, while Agent B, would be given later, say at $\tau, 0 < \tau < T$, time units after enrollment. As a result, the follow-up collected on Agent B, which ranges from τ to T, is less than the follow-up on Agent A, which ranges from 0 to T, and this differential follow-up might prove useful in the form of w_i used. Furthermore, there is no reason to assume that both agents have the same underlying distribution of DLT times, nor that these two distributions are independent.

Specifically, for the wgCRM, we can use non-decreasing functions $u(t_i)$ and $v(t_i)$, where $u(0) = v(0) = 0$ and $u(T) = v(T) = 1$, to denote the respective marginal contributions that Agents A and B make to the non-decreasing weight function $w[u(t_i), v(t_i)]$, where $w(0,0) = 0$, $w(1,1) = 1$. Note that we allow the possibility of a non-zero weight to any patient who has received Agent A and has yet to receive Agent B. This is because the time followed on Agent A alone still provides information about toxicity of the combined treatments, i.e if no DLT has yet occurred when receiving Agent A alone, this suggests that the combination of both agents may also prove to be non-toxic.

There are several possible forms for $w(\cdot)$, among which are three naive candidates:

$$w_1(t) = \min[u(t), v(t)] \tag{4.3}$$

$$w_2(t) = \max[u(t), v(t)] \tag{4.4}$$

$$w_3(t) = [u(t) + v(t)]/2 \tag{4.5}$$

Thus, if we were to define $u(t_i) = t_i/T$ and $v(t_i) = (t_i - \tau)/(T - \tau)$, then the three

weight functions above would be, respectively, the amount of follow-up on Agent B, the amount of follow-up on Agent A, and the average amount of follow-up on both agents.

The major limitation of these three weight functions is their lack of statistical motivation, as the appropriate weight should be based upon the joint distribution of the time-to-DLT for Agent A and the time-to-DLT for Agent B. Similar to the derivation made in *Braun* (2006) for the single-agent setting, let T_A and T_B denote the marginal underlying DLT times for Agent A and Agent B, with $T_A \in [0, T]$ and $T_B \in [\tau, T]$, and let T_F denote the observed failure time in $[0, T]$. Then, the appropriate weight is for a patient with follow-up t is

$$\begin{aligned} w(t) = Pr(T_F \leq t) &= Pr(\min(T_A, T_B) \leq t) \\ &= Pr(T_A \leq t \text{ or } T_B \leq t) \\ &= Pr(T_A \leq t) + Pr(T_B \leq t) - Pr(T_A \leq t, T_B \leq t) \\ &= u(t) + v(t) - C_\theta[u(t), v(t)] \end{aligned}$$

We now see that the appropriate forms for $u(t)$ and $v(t)$ are the respective marginal CDFs for T_A and T_B evaluated at t. Furthermore, since the joint CDF of T_A and T_B is harder to conceptualize than their marginal distributions, and prior data may even exist regarding the marginal CDFs, we express the joint CDF as a copula, which is a parametric function that "couples" the two marginal CDFs into a single joint CDF, with the parameter θ quantifying the association between T_A and T_B (*Nelson*, 1999).

Thus, the appropriate weight requires selection of a copula, for which a vast number of choices exist. We have chosen to use a Farlie-Gumbel-Morgenstern (FGM) (*Nelson*, 1999) that has the simple analytical form $C_\theta(x, y) = xy + \theta xy(1-x)(1-y)$, with

$\theta \in [-1, 1]$, $\theta = 0$ indicating independence and $0 < \theta \leq 1$ ($-1 \leq \theta < 0$) indicating positive (negative) correlation. Thus, if we were to extend the idea of marginal uniform distributions for T_A and T_B as used in the TITE-CRM, and were to also assume that T_A and T_B were independent so that $Pr(T_A \leq t, T_B \leq t) = Pr(T_A \leq t)Pr(T_B \leq t)$, the appropriate weight would be

$$\begin{aligned} w(t) &= \frac{t}{T} + max\left\{0, \frac{t-\tau}{T-\tau}\right\} - \left(\frac{t}{T}\right)\left(max\left\{0, \frac{t-\tau}{T-\tau}\right\}\right) \\ &= \frac{t}{T} + \left(1 - \frac{t}{T}\right)\left(max\left\{0, \frac{t-\tau}{T-\tau}\right\}\right), \end{aligned}$$

which has the simple interpretation as a weighted average of complete follow-up (weight of 1) and the amount of follow-up on Agent B, with the weight determined by the amount of follow-up on Agent A. In our simulations, we will examine using three weight functions, each of which assume marginal uniform DLT times and an FGM copula, with each of the functions assuming one of the values for $\theta \in \{-1, 0, 1\}$, which correspond to Spearman rank correlation values of -1/3, 0, and 1/3, respectively. Note that greater values of correlation between T_A and T_B are not possible, and this is one of the limitations of the FGM copula.

In an alternate use of the FGM copula, we see that a copula is a function that takes two arguments that lie between 0 and 1 and produce a value that also lies between 0 and 1. Thus, the copula itself could be a weight, i.e. $w(t) = Pr(T_A \leq t, T_B \leq t) = u(t)v(t)\{1 + \theta[1 - u(t)][1 - v(t)]\}$. Although such a weight function was not generated from statistical principals, it does have the property that for a given value of t, smaller (larger) values of θ lead to smaller (larger) weights. If late-onset DLTs were expected, then we would perhaps like to place less weight on a patient than if uniformly distributed DLTs were expected, as we would have less

confidence about how the patients' current follow-up without DLT would predict for later follow-up without DLT. Conversely, if early-onset DLTs were expected, then we would perhaps like to place more weight on a patient than if uniformly distributed DLTs were expected, as we would have more confidence about how the patients' current follow-up without DLT would predict for later follow-up without DLT. As a result, a value of $\theta < 1$ could be used if early-onset DLTs were expected and a value of $\theta > 1$ could be used if late-onset DLTs were expected. In our simulations, we will also examine this approach with values of $\theta \in \{-1, 0, 1\}$.

In summary, we have proposed a total of nine possible weight functions. The first three weights are given in Equations (4.3) - (4.5). The next three use the FGM copula with values of $\theta \in \{-1, 0, 1\}$:

$$w_4(t) = u(t)v(t)\{1 - [1 - u(t)][1 - v(t)]\} \tag{4.6}$$

$$w_5(t) = u(t)v(t) \tag{4.7}$$

$$w_6(t) = u(t)v(t)\{1 + [1 - u(t)][1 - v(t)]\} \tag{4.8}$$

and the final three are the theoretically motivated weights assuming an FGM copula with values of $\theta \in \{-1, 0, 1\}$:

$$w_7(t) = u(t_i) + v(t_i) - w_4(t) \tag{4.9}$$

$$w_8(t) = u(t_i) + v(t_i) - w_5(t) \tag{4.10}$$

$$w_9(t) = u(t_i) + v(t_i) - w_6(t) \tag{4.11}$$

For comparison purposes, we will also examine the use of a final weight function, $w_{10}(t) = 0$, which means that partially followed patients will be excluded entirely

from the likelihood, leading to use of only completely-followed subjects as proposed with the original CRM.

4.2.3 Trial Conduct

In Chapter 3, we implemented a stopping rule with the gCRM such that after at least three patients have been completely followed, the trial can be terminated if the lower bound of the 95% confidence interval is higher than the targeted rate Γ. However, during rapid accrual of patients, such a stopping rule would require enrollment of too many patients before trial termination would be considered. As an alternative, our stopping rule with the wgCRM will be similar to the one used in *Yuan and Yin* (2011a). Specifcally, when patient $(M+1)$ arrives, $1 \leq M \leq (N-1)$, for each of the posterior draws of the hyperparameters, we calculate an empirical probability of p_{11}, denoted as $\hat{p}_{11}$. If more than a fraction, ϕ, of the posterior draws have $\hat{p}_{11}$ greater than Γ, then we stop the trial and conclude all the dose combinations under study are overly toxic. i.e., a trial using the wgCRM should be terminated if $P(p_{11} > \Gamma) > \phi$. By using such a stopping rule, an overly toxic trial can be terminated very early, so as to enhance patient safety. Also, by varying ϕ, the investigator has the option to choose between conservative and aggressive stopping rules.

Like we did with the gCRM, we still limit the patient escalation or de-escalation to within the neighborhood of the current dose assignment as a trade-off between patient safety and full exploration of the dose combination space. The conduct of a trial with partial follow-up via the wgCRM is as follows:

(1) Elicit *a priori* estimates for the probabilities of DLT for combinations including the lowest dose of either agent;

(2) Determine appropriate values for the hyperparameters and rescaled dose values as described in Section 3.2.2, as well as the cut-off value ϕ for the stopping rule;

(3) Assign the first patient to combination $(1, 1)$;

(4) When patient $i = 2, 3, \ldots N$ arrives, compute a weight for any previously enrolled patient without DLT who is still under follow-up using the one of the weight functions described earlier;

(5) Determine if the stopping rule above has been met. If so, terminate the trial and conclude all the dose combinations under study are overly toxic; if not, go to step (6).

(6) Compute the posterior means $\tilde{p}_{jk}$ for each combination (j, k) using the method described in Section 4.2.2. If we let (j_{i-1}, k_{i-1}) denote the combination assigned to patient $i - 1$, we define the set of acceptable combinations as $S = \{(j, k) : \max(1, j_{i-1}-1) \leq j \leq \min(j_{i-1}+1, m), \max(1, k_{i-1}-1) \leq k \leq \min(k_{i-1}+1, n)\}$. The current estimate of the MTC, (j^*, k^*), is the combination in S whose value of $\tilde{p}_{jk}$ is closest to Γ;

(7) If data have been collected on all N patients, use step (6) to compute the final estimate of the MTC.

4.3 Simulation Studies

4.3.1 General Description

We now compare the performance of the ten weight functions described in Section 4.2.2 via simulation. We have a trial designed to study the sixteen combinations

of four doses of Agent A and four doses of Agent B. The goal is to determine the combination with a DLT rate of $\Gamma = 0.20$. A total of $N = 35$ patients will be enrolled in the trial. Each patient will be followed for $T = 100$ time units after enrollment. Agent A will be administered at enrollment, while Agent B will be administered at time $\tau = 20$ after enrollment. We assume that an average of 15 patients will arrive uniformly during any follow-up interval $[0, T]$, so that there is a considerable amount of partial follow-up that will exist in a trial. We also ran simulations with less than 15 subjects per follow-up interval, but found the differences among the weights are less distinguishable with less partial follow-up.

The true probabilities of DLT by time T for each combination are exactly the same as Scenarios A-G used in Section 3.3.1 with the gCRM. By doing so, we are able to use the results when using the gCRM with complete follow-up as a benchmark for the performance of each of the weight functions and determine how much information is lost with partial follow-up. All prior parameter values are the same as they were with the gCRM, and 3,000 draws were made from each posterior distribution, with a burn-in of the first 600 samples. A value of $\phi = 0.8$ was used for the threshold in the stopping rule.

We performed 1,500 simulations for each weight function in each scenario and each simulation was summarized by the combination selected as the MTC and the number of patients assigned to each combination.

4.3.2 When Weights are Based on Actual Distribution of DLT Times

In this section, the actual joint and marginal distributions of the DLT times for both agents were known and used directly in our methods. Specifically, for patients who experience DLT by time T, the actual time-to-DLT for Agent A was normal

with mean 70 and standard deviation 25, truncated to lie within the interval $[0, 100]$, while the actual time-to-DLT for Agent B was normal with mean 60 and standard deviation 20, truncated to lie within the interval $[20, 100]$. Thus, for a given follow-up $0 \leq t \leq 100$, we have

$$\begin{aligned} u(t) &= \frac{\Phi_{70,25}(t) - \Phi_{70,25}(0)}{\Phi_{70,25}(100) - \Phi_{70,25}(0)} \\ \\ v(t) &= \frac{\Phi_{60,20}(t) - \Phi_{60,20}(20)}{\Phi_{60,20}(100) - \Phi_{60,20}(20)}, \end{aligned}$$

where $\Phi_{m,s}(x)$ is the CDF at x for a normal distribution with mean m and standard deviation s. The two DLT times were also correlated using an FGM copula with $\theta = 1$. The observed DLT time for each patient with DLT was then selected as the minimum of the DLT times of Agents A and B. Thus, $w_9(t)$ given in Equation (4.11) is the theoretically "best" weight to use. The first four columns of Tables 4.1 - 4.7 display the simulation results.

From Table 4.1, we see that use of the optimal weight w_9 lead to similar operating characteristics as those when using weights w_2, w_7 and w_8, and these four weights have the best performance among all the weights, although the magnitude of the difference from the other weights is marginal, except w_10, which is clearly a poor choice. Specifically, the four best weights have the highest percentages of selecting an MTC around the targeted dose combination, with at least 54% of simulations identifying the MTC at combinations with DLT rates within four points of Γ, and at least 92% within ten points of Γ. They have the highest average percentages of patients assigned to the doses at or around the targeted dose combination, with about 34%-36% and 64%-68% of patients assigned to combinations with DLT rates within

four and ten points, respectively, of Γ. Similar findings are made from the results in Table 4.2.

Scenario C is a difficult setting for all of the weight functions, because the targeted dose combination is among one of the lowest dose combinations and the majority of the dose combinations are overly toxic, so that such a trial is has a high probability of being mistakenly terminated. From Table 4.3, we see that the optimal weight w_9 along with w_2, w_3, w_7, and w_8 terminate the trials in 7% or less of the simulations, while for the rest of the weights the trials are terminated in 8% or more of the simulations. All five weights identify the MTC within the five-point window of Γ in at least 52% of simulations and within the ten-point window of Γ in at least 84% of simulations.

Scenario D is a situation when all the dose combinations are overly toxic, and from Table 4.4 we see that, except for w_{10} which terminates the trials in all the simulations and the copula type weights w_4, w_5 and w_6 that terminate more than 93% of the trials, all the rest of the weights select the lowest dose combination as the MTC too often, and assigns too many patients to overly toxic dose combinations. However, we do not think such a scenario is very likely to happen in reality, since in applications, usually at least the lowest dose combination in a trial is chosen to be safe. It is also possible that the lower-than-desired rate of earlier terminations may be due to the stopping rule we selected and other stopping rules might lead to higher rates of early termination.

Tables 4.5-4.7 continue to confirm what was seen earlier: the optimal weight w_9 along with the weights w_1, w_2, w_7 and w_8 have better performance than the rest of the weight functions. Moreover, use of w_{10}, is a poor choice, and we can always do better to incorporate partial information than ignore it altogether.

4.3.3 When Weights are Based on Assumed Distribution of DLT Times

In Section 4.3.2, the true joint distribution of DLT times was known and was used to specify the forms of $u(t)$ and $v(t)$ as well as the theoretically appropriate form for $w(t)$. We now assess the performance of the ten weight functions when the true joint distribution and true marginal distributions of the two DLT times are *unknown*. Specifically, we use the functions $u(t) = t/T$ and $v(t) = (t-\tau)/(T-\tau)$, which assume uniform DLT times over $[0, T]$ for Agent A and $[\tau, T]$ for Agent B. In reality, DLTs are simulated to occur uniformly over the interval $[0, \tau]$, when Agent B has yet to be administered, with cumulative probabilities of DLT equal to $(0.01, 0.02, 0.03, 0.04)$ for doses 1, 2, 3, and 4 of Agent A, respectively. After Agent B has been administered, DLTs occur uniformly in the last third of the combined follow-up period, e.g. in the last $[(T-\tau)/3]$ time units of the follow-up period. All other parameters and values used in the simulations are the same as those in Section 4.3.2. The simulation results are displayed in the four columns of Tables 4.1 - 4.7.

From Table 4.1, we see that w_2, w_7, w_8 and w_9 have comparable performance that is better than the rest of the weights. Combinations with DLT rates within four points of the targeted DLT rate are identified as the MTC in 54%-55% of simulations when using one of these weights, as compared to no more than 50% with any of the other weights. Furthermore, combinations with DLT rates within ten points of the targeted DLT rate are identified at the MTC in 92%-95% of simulations when using one of these weights, as compared to no more than 91% with any of the other weights. This trend continues with the average percentage of patients assigned to each combination. Specifically, when using w_2, w_7, w_8 and w_9, 32%-36% (64%-68%) of patients are assigned to combinations with DLT rates within four (ten) points of

Γ, as compared to around 31% (58%) for the remaining weights. These four weights also lead to some of the lowest rates of early termination.

In Tables 4.2 and 4.3, we see that using weights w_2, w_3, which are the maximum and mean amounts of follow-up, respectively, and weights w_7, w_8, and w_9, which assume the DLT times are possibly correlated with uniform marginal CDFs, all lead to the best performance of the wgCRM, both in terms of identifying the MTC and assigning patients to combinations with DLT rates close to the targeted DLT rate. Within these five weights we see that w_3 does not perform as well as the other weights. Given that the DLT times are independent and that DLTs occur rather late (rather than uniformly), none of the weights is expected to be "best" and it is unclear why w_7, w_8, and w_9, which all make strong distributional assumptions about the DLT times, perform better than the other weights. Most notably, weights w_4, w_5, and w_6 all lead to unacceptably high rates of early termination. Tables 4.2 and 4.3 also demonstrate how inclusion of only partially-followed patients, using w_{10}, tends to identify the MTC less often than the other weights, and more importantly, assigns a much smaller fraction of patients to desirable combinations.

In Table 4.4, which summarizes the setting when all combinations are overly toxic, only the copula-type weights w_4, w_5 and w_6, as well as w_1 and w_{10} correctly terminate the trial in 91% or more of the trials, while all other weight functions lead to early termination rates of 77% or less. Use of these copula-type weights also leads to enrollment of fewer patients, on average. However, because use of the copula-type weights always tends to select combinations with low DLT rates as the MTC and higher rates of mistakenly terminating a trial, this is the only scenario where use of the copula-type weights leads to better operating characteristics than using any of the other weights. From Tables 4.5 - 4.7, we continue to see that using weights w_2,

w_7, w_8, and w_9 all lead to similar operating characteristics that are better than those when using the remaining weights.

4.4 Concluding Remarks

Based upon the results of our simulation study, we make the following general conclusions. First, incorporation of data from patients with incomplete follow-up is always better, in terms of identifying combinations close to the MTC and assignment of patients to those combinations, than basing decisions solely using data from patients with complete follow-up. Most notably, incorporation of incomplete follow-up allows for faster escalation of doses of either agent and thereby will always assign more patients to the MTC when it includes a high dose of either agent. Second, use of the FGM copula-type weights w_4, w_5 and w_6, as compared to the other weights, tends to cause over-estimation of the DLT rates of all combinations, which results in assignment of more patients to dose combinations whose DLT rates are less than the desired DLT rate Γ and also higher rates of early termination. However, we chose the FGM family of copulas mainly because of its simple analytical form and perhaps other copulas may produce weights that result in better operating characteristics.

Third, as theoretically shown, our simulations also support using weights that are based upon the true joint distribution of the time-to-DLT of both agents. Doing so leads to a higher percentage of correct identifying the MTC, the average percentage of patients assigned to combinations with DLT rates close to that desired, and a lower rate of early termination in trials in which at least one combination is acceptable. However, the optimality of weight w_7, w_8, or w_9, depending upon the amount of correlation between the DLT times of the agents, is impacted by the incorporation of

a stopping rule. Specifically, in scenarios where many of the dose combinations have DLT rates above the target Γ and the MTC exists at a combination with a low dose of either agent, use of these weights leads to higher rates of early termination, which in turn lessens the ability of the wgCRM to correctly identify the MTC.

Fourth, even when the joint distribution of the time-to-DLT of both agents is known, use of the "naive" weights w_2, the larger of the two marginal CDFs, and w_3, the average of the marginal CDFs, tends to produce operating characteristics that are comparable to those when using the "optimal" weights. Fifth, when the joint distribution of the DLT times of the agents is not known, we continued to see comparable operating characteristics among the weights w_2, w_3, w_7, w_8, and w_9 when assuming uniform margins for the times to DLT of each agent. However, our simulations focused upon one specific example of actual DLT times and our results may not be generalizable to other settings with different marginal failure time distributions of the two agents.

Table 4.1: Operating characteristics of the wgCRM in Scenario A using 10 different weights. All values are multiplied by 100.

	Known Joint & Marginal Distribution of DLT Times					Unknown Joint & Marginal Distribution of DLT Times				
	Within 4 pts	Within 10 pts	Lower than 10 pts	Higher than 10 pts	Early Termination	Within 4 pts	Within 10 pts	Lower than 10 pts	Higher than 10 pts	Early Termination
	Average Percentage of MTC Selections									
gCRM	59	90	2	6	2	59	90	2	6	2
w_1	50	91	3	4	2	40	75	6	2	17
w_2	54	92	1	4	3	46	90	6	1	3
w_3	48	88	4	5	3	44	86	7	2	5
w_4	43	84	5	4	7	32	60	3	2	35
w_5	48	88	4	5	3	32	65	4	1	30
w_6	49	89	4	4	3	35	70	4	1	25
w_7	54	92	2	4	2	49	91	6	2	1
w_8	54	95	1	3	1	49	92	5	1	2
w_9	55	95	1	4	0	51	92	5	2	1
w_{10}	38	68	3	5	24	49	86	5	5	4
	Average Percentage of Patient Assignments									
gCRM	45	77	10	13	0	45	77	10	13	0
w_1	33	66	10	13	0	25	57	28	3	12
w_2	36	68	6	24	2	36	75	19	7	0
w_3	29	58	31	6	5	32	71	23	4	2
w_4	28	57	34	5	4	17	43	31	2	24
w_5	31	58	31	6	5	19	46	31	1	22
w_6	30	62	29	6	3	23	51	30	2	17
w_7	34	66	5	28	1	38	77	14	8	1
w_8	32	64	5	28	3	37	77	16	7	0
w_9	34	65	5	28	2	37	76	15	7	2
w_{10}	14	34	48	1	17	20	45	52	4	0

Table 4.2: Operating characteristics of the wgCRM in Scenario B using 10 different weights. All values are multiplied by 100.

	Known Joint & Marginal Distribution of DLT Times				Unknown Joint & Marginal Distribution of DLT Times			
	Within 5 pts	Within 10 pts	Lower than 10 pts	Early Termination	Within 5 pts	Within 10 pts	Lower than 10 pts	Early Termination
	Average Percentage of MTC Selections							
gCRM	80	95	5	0	80	95	5	0
w_1	65	93	6	1	37	67	20	13
w_2	76	94	5	1	43	80	19	1
w_3	70	93	5	2	39	75	22	3
w_4	57	85	13	2	31	54	19	27
w_5	60	86	12	2	31	60	22	18
w_6	62	89	10	1	34	63	19	18
w_7	79	96	2	2	48	82	18	0
w_8	78	95	3	2	49	83	18	0
w_9	76	95	4	1	47	81	18	1
w_{10}	47	74	14	12	46	79	18	3
	Average Percentage of Patient Assignments							
gCRM	66	84	17	0	66	84	17	0
w_1	39	60	40	0	17	37	57	6
w_2	64	83	17	0	32	58	40	2
w_3	55	77	22	1	25	49	50	1
w_4	22	44	53	3	10	24	58	18
w_5	25	46	51	3	11	26	59	15
w_6	28	49	51	0	12	29	57	14
w_7	71	85	12	3	39	63	37	0
w_8	70	84	13	3	37	63	37	0
w_9	68	85	13	2	37	63	37	1
w_{10}	12	29	63	8	17	34	61	5

Table 4.3: Operating characteristics of the wgCRM in Scenario C using 10 different weights. All values are multiplied by 100.

	Known Joint & Marginal Distribution of DLT Times				Unknown Joint & Marginal Distribution of DLT Times			
	Within 5 pts	Within 10 pts	Higher than 10 pts	Early Termination	Within 5 pts	Within 10 pts	Higher than 10 pts	Early Termination
	Average Percentage of MTC Selections							
gCRM	52	86	6	8	52	86	6	8
w_1	48	82	10	8	35	55	7	38
w_2	52	86	9	5	49	80	8	12
w_3	52	84	9	7	46	76	8	16
w_4	43	69	10	21	21	31	5	64
w_5	44	73	11	16	27	41	6	53
w_6	47	77	11	12	29	44	5	51
w_7	54	88	9	3	53	86	9	5
w_8	57	89	9	2	51	85	9	6
w_9	54	87	9	4	50	83	9	8
w_{10}	25	40	7	53	43	70	14	16
	Average Percentage of Patient Assignments							
gCRM	37	75	21	4	37	75	21	4
w_1	31	63	31	6	25	55	19	26
w_2	28	49	49	2	32	60	32	8
w_3	32	55	42	3	33	65	27	8
w_4	26	66	18	16	17	44	8	48
w_5	28	67	22	11	20	49	11	40
w_6	30	68	25	7	20	49	10	41
w_7	27	44	55	1	33	59	39	2
w_8	26	43	56	1	33	59	39	2
w_9	26	44	49	7	32	59	38	3
w_{10}	9	54	8	38	13	68	26	6

Table 4.4: Operating characteristics of the wgCRM in Scenario D using 10 different weights. All values are multiplied by 100.

	Known Joint & Marginal Distribution of DLT Times			Unknown Joint & Marginal Distribution of DLT Times		
	Lowest Dose	Other Doses	Early Termination	Lowest Dose	Other Doses	Early Termination
	Average Percentage of MTC Selections					
gCRM	5	0	95	5	0	95
w_1	14	1	85	7	2	91
w_2	39	10	51	27	6	67
w_3	33	6	61	19	4	77
w_4	5	0	95	2	0	98
w_5	6	0	94	3	0	97
w_6	7	0	93	3	0	97
w_7	48	13	39	35	9	66
w_8	47	13	40	35	6	59
w_9	47	12	41	31	8	61
w_{10}	0	0	100	9	0	91
	Average Percentage of Patient Assignments					
gCRM	31	9	60	31	9	60
w_1	31	19	50	22	11	67
w_2	26	52	22	27	45	28
w_3	28	48	24	26	34	40
w_4	29	7	64	18	4	78
w_5	28	9	63	18	7	75
w_6	28	9	63	19	9	72
w_7	24	61	15	26	50	24
w_8	24	62	14	28	48	24
w_9	23	62	15	27	47	26
w_{10}	26	0	74	50	7	43

Table 4.5: Operating characteristics of the wgCRM in Scenario E using 10 different weights. All values are multiplied by 100.

	Known Joint & Marginal Distribution of DLT Times					Unknown Joint & Marginal Distribution of DLT Times				
	Within 2 pts	Within 10 pts	Lower than 10 pts	Higher than 10 pts	Early Termination	Within 2 pts	Within 10 pts	Lower than 10 pts	Higher than 10 pts	Early Termination
	Average Percentage of MTC Selections									
gCRM	42	74	11	12	13	42	74	11	12	3
w_1	28	67	12	17	4	25	49	11	11	29
w_2	31	72	5	19	4	32	68	14	14	4
w_3	31	71	8	18	3	34	65	14	9	12
w_4	27	62	13	11	14	16	34	7	9	50
w_5	29	61	14	14	11	19	39	7	8	46
w_6	27	63	16	15	6	21	41	10	9	40
w_7	33	74	5	19	2	33	69	14	12	5
w_8	32	72	6	20	2	33	68	15	14	3
w_9	31	73	6	19	2	31	69	14	13	4
w_{10}	17	35	9	12	44	28	62	14	14	10
	Average Percentage of Patient Assignments									
gCRM	29	59	19	21	1	29	59	19	21	1
w_1	27	53	29	16	2	21	38	32	8	22
w_2	24	51	10	35	4	30	58	25	16	1
w_3	27	57	13	28	2	30	51	28	10	11
w_4	23	41	40	8	11	17	28	31	5	36
w_5	24	46	38	11	5	18	31	30	6	33
w_6	26	47	36	11	6	19	34	32	6	28
w_7	22	48	8	43	1	30	58	21	20	1
w_8	22	47	8	41	4	31	58	22	18	2
w_9	23	51	9	40	0	30	59	21	18	2
w_{10}	12	19	48	5	28	17	30	56	9	5

Table 4.6: Operating characteristics of the wgCRM in Scenario F using 10 different weights. All values are multiplied by 100.

	Known Joint & Marginal Distribution of DLT Times				Unknown Joint & Marginal Distribution of DLT Times			
	Within 4 pts	Within 10 pts	Higher than 10 pts	Early Termination	Within 4 pts	Within 10 pts	Higher than 10 pts	Early Termination
	Average Percentage of MTC Selections							
gCRM	44	84	9	7	44	84	9	7
w_1	33	80	11	9	25	60	6	34
w_2	42	81	17	2	33	82	11	7
w_3	40	80	16	4	30	77	9	14
w_4	26	69	8	23	13	35	5	60
w_5	28	71	10	19	18	41	6	53
w_6	30	76	9	15	20	49	6	45
w_7	46	81	18	1	37	84	12	4
w_8	42	77	20	3	36	84	12	4
w_9	45	79	17	4	38	82	11	7
w_{10}	40	80	16	4	30	77	9	14
	Average Percentage of Patient Assignments							
gCRM	40	81	17	2	40	81	17	2
w_1	29	72	23	5	21	63	12	25
w_2	31	54	43	3	32	71	24	5
w_3	34	60	36	4	29	72	18	10
w_4	20	72	13	15	13	50	5	45
w_5	21	73	15	12	16	52	7	41
w_6	23	73	16	11	18	59	11	30
w_7	31	51	49	0	33	68	29	3
w_8	31	49	49	2	33	69	28	3
w_9	31	50	49	1	32	68	29	3
w_{10}	6	54	4	42	13	73	16	11

Table 4.7: Operating characteristics of the wgCRM in Scenario G using 10 different weights. All values are multiplied by 100.

	Known Joint & Marginal Distribution of DLT Times					Unknown Joint & Marginal Distribution of DLT Times				
	Within 5 pts	Within 10 pts	Lower than 10 pts	Higher than 10 pts	Early Termination	Within 5 pts	Within 10 pts	Lower than 10 pts	Higher than 10 pts	Early Termination
	Average Percentage of MTC Selections									
gCRM	59	91	5	4	0	59	91	5	4	0
w_1	48	77	22	3	0	38	67	18	4	11
w_2	50	74	22	1	3	41	71	24	3	2
w_3	48	74	21	1	4	39	70	23	5	2
w_4	45	74	17	6	3	33	60	13	6	21
w_5	46	75	19	4	2	32	60	16	8	16
w_6	46	74	21	4	1	32	63	16	7	14
w_7	49	71	26	0	3	43	73	26	1	0
w_8	49	73	22	1	4	43	73	26	1	0
w_9	47	71	26	1	2	41	70	27	2	1
w_{10}	33	64	17	13	6	35	66	28	5	1
	Average Percentage of Patient Assignments									
gCRM	38	75	18	8	0	38	75	18	8	0
w_1	23	46	33	20	1	18	38	46	9	7
w_2	25	48	25	23	4	24	48	35	15	2
w_3	27	51	26	21	2	21	45	37	14	4
w_4	20	41	45	13	1	13	28	50	6	16
w_5	21	43	42	14	1	14	32	51	7	10
w_6	22	44	41	16	0	16	34	52	7	7
w_7	27	50	22	26	2	25	50	30	20	0
w_8	28	51	22	26	1	25	50	30	20	0
w_9	26	50	22	26	2	25	52	28	19	1
w_{10}	11	29	58	8	5	10	26	57	16	1

CHAPTER V

Summary and Future Work

This dissertation focuses on two outstanding issues in phase I oncology trials: one is how to incorporate the partial information from patients that are still under follow-up in order to instantly assign a treatment to the next available patient right at his arrival; the other issue is approaching efficient methods with higher accuracy, less required information, less complexity, and more robustness to model mis-specifications to handle dose-finding in dose combinations from two agents.

In Chapter II and Chapter IV, we developed methods that dealt with incomplete follow-up in either one-agent or two-agent trials. Via simulations we showed that our proposed methods had the ability to reduce the overall duration of trial while consistently maintaining high accuracy in targeting the correct MTC or MTD, and assigning a high percentage of the patients to or around the recommended dose or dose combination. In Chapter III, we showed how to develop a generalized CRM using the logistic or power model of the CRM that extends the one-agent CRM to a two-dimensional space. We showed via simulations that, as compared to several other methods that aim to solve the same type of issue, the proposed method gCRM is less complex, more robust, and performs better in terms of both the frequency of

recommending the correct MTC or the frequency of assigning patients around the recommended dose combination.

The main idea used to generate the time-to-event versions of phase I trials is to incorporate adaptive weight functions into the algorithm. In Chapter II, we developed the time-to-event version of the BCD, and the weight function is calculated from the DLT outcomes of the previously followed patients, as well as the proportion of time that had been followed on patients that are still under follow-up. Besides the CRM and BCD, which are representatives of the parametric and non-parametric phase I dose-finding algorithms, and already have their time-to-event versions of the algorithm, there is another distinct type of Phase I trial, called the group up-and-down designs. In such designs, in stead of being enrolled one after another, the patients enter a trial in groups with at least two patients per group, and the same group of patients are assigned to the same dose at the same time. One future work is to develop time-to-event versions of such group up-and-down designs.

The gCRM proposed in Chapter III is a serious contender to several other two-agent phase I dose-finding algorithms in literature. However, in applications, particularly in the pharmaceutical area, usually patients are entered into a study in cohorts rather than one by one. Therefore, one future direction in the two-agent phase I trials is to develop methods that generalizes the group up-and-down designs from the one-dimensional space to the two-dimensional space.

The weighted gCRM proposed in Chapter IV generalizes the gCRM by allowing partial follow-up in the dose assignment decision rule. In the optimal weight we had proposed, we used the FGM family of Copulas. It should be noted that we chose FGM Copula for illustration purpose mainly because of its simple analytical form. The FGM Copula requires that the correlation between the failure times of the two

agents is known to the investigators. However, this correlation information may not always be available. On the other hand, there can be many other Copulas that can model the joint distribution of the failure times of the two agents, but require different information about the two agents other than their correlation. Due to the different requirements of various types of Copulas, investigators may choose other types of Copulas based on the available information from the two agents.

Overall speaking, this dissertation focuses on the applied aspect of statistics. It addresses practical issues in phase I oncology trials, and the methods developed here are easy to carry out, and are directly applicable in real life scenarios. One more step that will make the current work better is to develop software support. For example, building R packages, SAS macros, SPSS modules so that these algorithms can be more easily implemented.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahn, C. (1998), An evaluation of phase I cancer clinical trial designs, *Statistics in Medicine*, *17*, 1537–1549.

Bekele, B., Y. Ji, Y. Shen, and P. Thall (2008), Monitoring late-onset toxicities in phase I trials using predicted risks, *Biostatistics*, *9*, 442–457.

Braun, T. (2006), Generalizing the TITE-CRM to Adapt for Early- and Late-Onset Toxicities, *Statistics in Medicine*, *25*, 2071–2083.

Braun, T., and S. Wang (2010), A Hierarchical Bayesian design for phase I trials of novel combinations of cancer therapeutic agents, *Biometrics*, *66*, 805–812.

Cheung, Y., and R. Chappell (2000), Sequential designs for phase I clinical trials with late-onset toxicities, *Biometrics*, *56*, 1177–1182.

Conaway, M., S. Dunbar, and S. Peddada (2004), Designs for single or multiple-agent phase I trials, *Biometrics*, *60*, 661–669.

Dempster, A., N. Laid, and D. Rubin (1977), Maximum Likelihood From Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society*, *39*, 1–38.

Durham, S., and N. Flournoy (1995a), Up-and-down designs I: stationary treatment distributions, *IMS Lecture Notes - Monograph Series*, *25*, 139–157.

Durham, S., and N. Flournoy (1995b), Up-and-down designs II: exact treatment moments, *IMS Lecture Notes - Monograph Series*, *25*, 158–178.

Faries, D. (1994), Practical modifications of the continual reassessment method for phase I cancer clinical trials, *Journal of Biopharmaceutical Statistics*, *4*, 147–164.

Giovagnoli, A., and N. Pintacuda (1998), Properties of frequency distributions induced by general 'up-and-down' methods for estimating quantiles, *Journal of Statistical Planning and Inference*, *74*, 51–63.

Goodman, S., M. Zahurak, and S. Piantadosi (1995), Some practical improvements in the continual reassessment method for phase I studies, *Statistics in Medicine*, *14*, 1149–1161.

Korn, E., D. Midthune, T. Chen, L. Rubinstein, M. Christian, and R. Simon (1994), A comparison of two phase I trial designs, *Statistics in Medicine*, *13*, 1799–1806.

Nelson, R. (1999), *An introduction to copulas*, Springer, New York.

Onar-Thomas, A., and Z. Xiong (2010), A simulation-based comparison of the traditional method, Rolling-6 design and a frequentist version of the continual reassessment method with special attention to trial duration in pediatric Phase I oncology trials, *Contemporary Clinical Trials*, *31*, 259–270.

O'Quigley, J., and S. Zohar (2006), Experimental designs for phase I and phase I/II dose-finding studies, *British Journal of Cancer*, *94*, 609–613.

O'Quigley, J., M. Pepe, and L. Fisher (1990), Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer, *Biometrics*, *46*, 33–48.

Rowinsky, E., et al. (1996), Sequences of topotecan and cisplatin: Phase I, pharmacologic, and in vitro studies to examine sequence dependence, *Journal of Clinical Oncologys*, *14*, 3074–3084.

Shen, L., and J. O'Quigley (1996), Consistency of continual reassessment method under model misspecification, *Biometrika*, *83*, 395–405.

Skolnik, J., J. Barrett, and J. B (2008), Shortening the timeline of pediatric phase I trials: the rolling six design, *Journal of Clinical Oncology*, *26*, 190–195.

Storer, B. (1989), Design and analysis of phase I clinical trial, *Biometrics*, *45*, 925–937.

Storer, B. (2001), An evaluation of phase I clinical trial designs in the continuous dose-response setting, *Statistics in Medicine*, *20*, 2399–2408.

Stylianou, M., and N. Flournoy (2002), Dose finding using the biased coin up-and-down design and isotonic regression, *Biometrics*, *20*, 171–177.

Stylianou, M., and D. Follmann (2004), The accelerated biased coin up-and-down design in phase I trials, *Journal of Biopharmaceutical Statistics*, *14*, 249–260.

Tanner, M., and H. Wong (1987), The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, *82*, 528–540.

Thall, P., R. Millikan, P. Mueller, and S. Lee (2003), Dosefinding with two agents in Phase I oncology trials, *Biometrics*, *59*, 487–496.

Wages, N., M. Conaway, and J. O'Quigley (2011a), Continual reassessment method for partial ordering, *Biometrics*, *67*, 1555–1563.

Wages, N., M. Conaway, and J. O'Quigley (2011b), Dose-finding design for multi-drug combinations, *Clinical Trials*, *8*.

Wang, K., and A. Ivanova (2005), Two-dimensional dose finding in discrete dose space, *Biometrics*, *61*, 217–222.

Yin, G., and Y. Yuan (2009a), A latent contingency table approach to dose finding for combinations of two agents, *Biometrics*, *65*, 866–875.

Yin, G., and Y. Yuan (2009b), Bayesian dose finding in oncology for drug combinations by copula regression, *Journal of the Royal Statistical Society*, *58*, 211–224.

Yuan, Y., and G. Yin (2008), Sequential continual reassessment method for two-dimensional dose finding, *Statistics in Medicine*, *27*, 5664–5678.

Yuan, Y., and G. Yin (2011a), Baysian Phase I/II Adaptively Randomized Oncology Trials with Combined Drugs, *Annals of Applied Statistics*, *5*, 924–942.

Yuan, Y., and G. Yin (2011b), Robust EM Continual Reassessment Method in Oncology Dose Finding, *Journal of the American Statistical Association*, *106*, 818–831.