Georgia State University
ScholarWorks @ Georgia State University

Mathematics Dissertations | Department of Mathematics and Statistics

Spring 5-13-2013

Advanced Designs of Cancer Phase I and Phase II Clinical Trials

Ye Cui

Follow this and additional works at: https://scholarworks.gsu.edu/math_diss

Recommended Citation
Cui, Ye, "Advanced Designs of Cancer Phase I and Phase II Clinical Trials." Dissertation, Georgia State University, 2013.
https://scholarworks.gsu.edu/math_diss/15

This Dissertation is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ADVANCED DESIGNS OF CANCER PHASE I AND PHASE II CLINICAL TRIALS

by

YE CUI

Under the Direction of Ruiyan Luo and Zhengjia Chen

ABSTRACT

The clinical trial is the most import study for the development of successful novel drugs. The aim of this dissertation is to develop innovative statistical methods to overcome the three main obstacles in clinical trials: (1) lengthy trial duration and inaccurate maximum tolerated dose (MTD) in phase I trials; (2) heterogeneity in drug effect when patients are given the same prescription and same dose; and (3) high failure rates of expensive phase III confirmatory trials due to the discrepancy in the endpoints adopted in phase II and III trials. Towards overcoming the first obstacle, we originally develop a hybrid design for the time-to-event dose escalation method with overdose control using a normalized equivalent toxicity score (NETS) system. This hybrid design can substantially reduce sample size, shorten study length, and estimate accurate MTD by employing a parametric model and adaptive Bayesian approach. Toward overcoming the second obstacle, we propose a new approach to incorporate patients' characteristic using our

proposed design in phase I clinical trials which considers the personalized information for patients who participant in the trials. To conquer the third obstacle, we propose a novel two-stage screening design for phase II trials whereby the endpoint of percent change in of tumor size is used in an initial screening to select potentially effective agents within a short time interval followed by a second screening stage where progression free survival is estimated to confirm the efficacy of agents. These research projects will substantially benefit both cancer patients and researchers by improving clinical trial efficiency and reducing cost and trial duration. Moreover, they are of great practical meaning since cancer medicine development is of paramount importance to human health care.

INDEX WORDS: Clinical trial design, Phase I, Phase II, Maximum tolerated dose (MTD), Adaptive design, Personalized MTD, Covariate effect, Two-stage design, Success rate, Trial efficiency

ADVANCED DESIGNS OF CANCER PHASE I AND PHASE II CLINICAL TRIALS

by

YE CUI

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2013

Copyright by

Ye Cui

2013

ADVANCED DESIGNS OF CANCER PHASE I AND PHASE II CLINICAL TRIALS

by

YE CUI

Committee Chair: Ruiyan Luo

Committee: Zhengjia Chen

Xin Qi

Yichuan Zhao

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2013

DEDICATION

I dedicate this dissertation to my husband, who supported me each step of the way.

Also, my unreserved gratitude goes to my parents who always encouraged and support all of my scholarly endeavors.

ACKNOWLEDGEMENTS

The writing of this dissertation has been one of the most significant academic challenges I have ever had to face. I would not be able to finish this dissertation without the help of so many people in so many ways.

I would like to express my deepest appreciation to my committee chair, Dr. Ruiyan Luo for her guidance, understanding, patience, and most importantly, her friendship during my graduate studies at Georgia State. She was always there, listening and encouraging me for both my study and life.

I would like to gratefully and sincerely thank my co-advisor Dr. Zhengjia Chen, who continually and convincingly conveyed a spirit of adventure and an excitement in regard to research. Had I not attended his presentation, I would not have pursued the study of clinical trial design.

I would like to thank my committee members, Dr. Zhao and Dr. Qi, who patiently corrected my writing and gave me many valuable suggestions with my research and further work. I would also like to thank Dr. Qin, and Dr. Xiao, for guiding my research for the past several years and helping me to develop my background in statistics, and public health.

I would also thank Department of Mathematics and Statistics at Georgia State, especially those members of my doctoral committee for their input, valuable discussions and accessibility. My research would not have been possible without their helps.

Finally, and most importantly, I would like to thank my parents, and my family. They were always supporting me and encouraging me with their best wishes. And a special thank you to my husband, Zhibo Wang. He was always there cheering me up and stood by me through the good times and bad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

This dissertation consists of a series of studies conducted as part of cancer phase I and phase II clinical trial designs. The first part of this dissertation sought to determine how and to what extent the proposed hybrid phase I design reduces the trial duration and improve the dosage accuracy. The second part of the dissertation sought to adjust the current phase I design with respect to patients' individualized characteristic among the population. The third part of the dissertation sought to measure the improvement to the success rate and model sensitivity of proposed two-stage phase II trial design.

1.1 Statement of Need

Healthcare is a global topic. It is the common expectation for people to live a prolonged and high quality of life, regardless of where we inhabit, what we eat or how we live. In one way or another, people are connected by many of the same health issues, such as deaths from non-communicable diseases — cancer, cardiovascular disease, diabetes and so forth. Today, the purpose of medical research thus is to turn biomedical discoveries into practice to improve human health. It worth unremitting study via the help of clinical trial researches and is of paramount importance for the society as a whole.

Medical research has provided benefits for everyone with a longer and healthier life in the United States — a world leader in the field. However, challenges are still exists. Tremendous efforts have been made towards life-threatening diseases, such as cancer and AIDS on decades; but little has been known to these persistent unsolved problems. Hence, continuous studies are indispensable. Now, it has been the essential objectives of medical research to discover curative treatment for the life-threatening diseases, such as cancer.

Throughout the United States, cancer is the second leading cause of death, exceeded only by heart disease, accounting for nearly one in every four deaths (Siegel, et al., 2012). In 2012, about 1,638,910 new cancer cases would be diagnosed, and 577,190 Americans are expected to die of this disease, more than 1,500 people per day (ACS, 2012). Cancer causes severe pain and suffering for patients, as well as their families and friends. Moreover, the financial costs of cancer are high for both cancer patients and the society as a whole. According to National Institutes of Health estimates, direct cancer care costs $124.6 billion in the United States in 2010. It is expected that the expenditures will increase at a faster rate as cancer prevalence increases. Total costs will double if indirect costs are included, such as expenditure of cancer-related illness and death (NIH, 2012). Today, forty years after President Richard Nixon signed the National Cancer Act in 1971, with unremitting effort from all cancer researchers, the five-year relative survival rate — an important standard to compare the effectiveness of treatments for all cancers diagnosed between 2001 and 2007 grew to 67%, from 49% in the period 1975-1977 (Siegel, et al., 2012). This improvement in survival reflects both progresses in diagnosing certain cancers at an earlier stage and advance in medical treatment.

On the way of seeking effective therapeutic regimens for cancer, researchers have taken big step forward in both diagnosing and treating the disease by using mechanisms of advanced methods. Notably improvements in cancer treatment and relative survival rate have been made through the help of clinical trials since 1975 (Jemal, et al., 2008), which is the prevailing approach to develop candidate therapies.

1.2 Background Introduction

Cancer clinical trials are investigations that conducted in human beings intended to evaluate the safety and effects of new anti-cancer drug or drug combinations. Requirements to

regulate the conduction of such trials are provided by FDA nationally, or international organizations, such as EU.

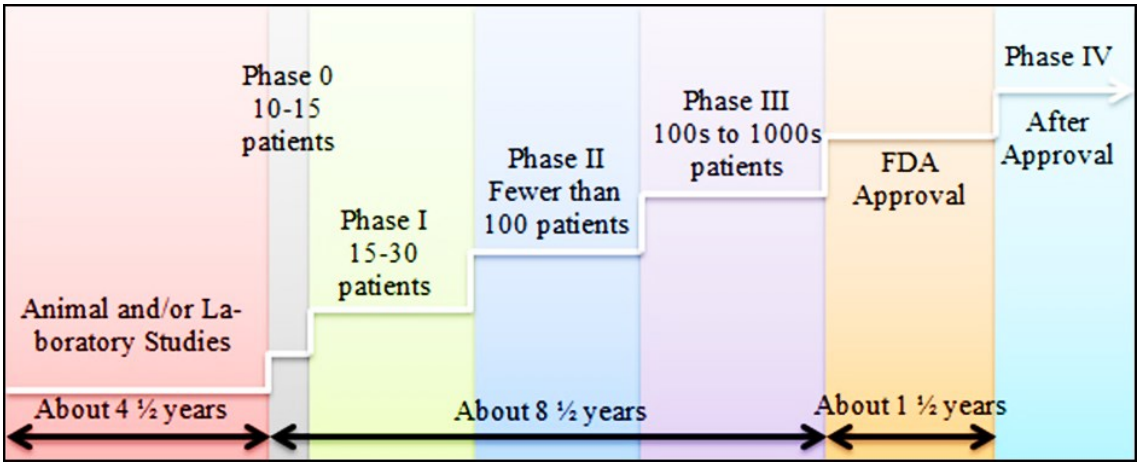

Figure 1-1 Cancer Clinical Trial Process

These studies apply the gold standard conveying biomedical theory and laboratory result into practical therapeutic regimens. There are several steps—called "phases" that conducted consecutively, with approval before performing each phase. Figure 1-1 illustrates the process of a complete cancer trial which lasts more than 10 years. Patients who participate in a cancer clinical trial, take part in only one phase most of the time. Phases are designed according to different purposes.

Phase 0 trials are conducted before initiating traditional phase I dose escalation, which primary goal is to evaluate pharmacodynamics and/or pharmacokinetic properties of the new drug or the drug combination (Gupta, et al., 2011). In response to Investigational New Drug (IND) guidance of FDA and European Medicines Agency, a phase 0 trial is an exploratory study (Robinson, 2008). It is conducted only on a small number of patients (approximately 10-15) by very small doses of the new drug or drug combinations as the safety or effectiveness of the new drug or the drug combinations is unknown. It is used to investigate whether the drug approaches the tumor, how the drug affects human body, and how cancer cells in the human body respond to

the drug. A significant difference between phase 0 trials and later phases of cancer trials is that phase 0 trials have no therapeutic intent (Kummar, et al., 2008). Because drug doses are low, there's also less risk to the patient in phase 0 studies compared to phase I studies. The patients who take part in these trials usually have advanced disease, and no known effective treatment options are available to them. Phase 0 studies help researchers find out whether the drugs do what they are expected to do. This process may help avoid the delay and expense of finding out years later in phase II or even phase III clinical trials that the drug doesn't act as it was expected to base on lab studies.

The principal purpose of cancer phase I clinical trials is to evaluate the safety of the new anticancer drug, and determine the recommended dose and/or schedule of an experimental drug or drug combination for efficacy testing in subsequent phase II trials (Lin, et al., 2001). A cancer phase I trial may involve new drugs or drug combinations under development or new combinations of already available drugs. It is usually conducted to small numbers of people (15 to 30), who are at advanced cancer stage that have progressed in spite of standard treatments (Tourneau, et al., 2009). Since little is known with the new drugs or the drug combinations under testing, it is essential to conduct a phase I trial carefully and thoughtfully. Usually the study begins at a low dose and move forward if the treatment is considered to be safe. A distinction between cancer phase I trials and other phase I trials is that the former ones have therapeutic aim. The primary assumption underlying an anticancer drug is that toxicity is a prerequisite for optimal antitumor activity. Therefore, cancer phase I designs should minimize exposures to unnecessary non-therapeutic doses while preserve safety by minimize the number of patients given severely toxic overdoses (Babb, et al., 1998). Investigation result — optimal recommended

dose level or so called maximum tolerated dose (MTD) will be recommended in the subsequent phase II and phase III trials.

Phase II studies play a pivotal role in anticancer drug development, which are conducted to assess whether a new drug or a drug combination has sufficient the efficacy to warrant further investigation. A cancer phase II trial usually involves fewer than 100 patients who have a specific type of cancer or related cancers. Adopting information of MTD from preceding phase I trials, these phase II studies aim to obtain a preliminary estimate of the effectiveness of antitumor activities of new drugs or drug combinations (Brown, et al., 2011). Patients who take part in phase II trials may or may not have been treated previously with standard treatment. In cancer phase II trials effectiveness is usually measured by tumor response (shrinkage) that a 50% or more shrinkage in tumor size is considered to be a successful treatment. Various alternate endpoints have been proposed to assess the efficacy of treatment such as progression-free survival (PFS), overall survival, biomarkers, and more recently, evaluation of tumor size as a continuous variable (Dhani, et al., 2009). New drugs or drug combinations showing sufficient activity in the phase II setting might be evaluated subsequently in a phase III comparative trial.

Phase III trials are the last step before approval, which primary goal is to compare the effectiveness and side effects of the new drugs or drug combinations with the current standard treatment for specific type of cancer. Phase III trials take one step forward than phase II trials that only assess the effectiveness without comparing with standard treatment. If a new drug or a drug combination is proven to be effective but shows no significant benefits over the current used treatment; it is still considered as “failed”, and will not be approved for production. If a new drug or a drug combination is more effective than the standard treatment and/or is easier to tolerate, it may become the new standard of treatment. A large number of patients (100s to

1000s) are included in a phase III trial. A phase III trial usually designs as a randomized two arms experiment; because patients who participate in phase III trials may or may not have been treated previously as in proceeding phase II trials. New drugs or drug combinations that identified as "success" in phase III studies will be approved for market and also be observed for post-marketing surveillance trials to follow up safety and effectiveness issues (NCI, 2013). In Table 1.1 we summarize the characteristics of phase I through phase III trials in the process of anticancer drug development.

Table 1.1Characteristic of cancer phase I, phase II, and phase III trials.

Category	Phase I	Phase II	Phase III
Objective	Estimate MTD, determine toxicity spectrum, and evaluate pharmacokinetics.	Evaluate new drug or treatment effectiveness.	Compare with standard drug or treatment, assess efficacy (clinical benefit).
Cancer Type	All types of cancer.	Specific type of cancer, or related cancers.	Specific type of cancer.
Dose	Escalated.	MTD	MTD
Endpoint	Toxicity.	Tumor response, event-free survival.	Overall survival.
Design	Dose escalation in small cohorts of patients.	Two-stage or multistage design (early stopping rule).	Randomized design ($\pm$ blinded).

A phase IV trials — also called a post-marketing surveillance trial is the stage after a new drug or drug combination has been approved by the FDA or other regulatory agencies for standard use. The purpose of cancer phase IV studies is to continue investigating the new drugs or drug combinations with the treatment effects and long-term safety. Phase IV trials usually involves large number of patients from several hundred to as many as several thousand people. These studies play a more and more important role in oncology drug development as it is necessary to continual assess new drugs or drug combinations base on large population. Finally,

investigators will determine further research plans with the data collected during the trials (NCI, 2013).

1.3 Purpose of the Study

In developing new anti-cancer drugs, statistics is an indispensable and crucial element since clinical trial is a process that applies statistical inference theory into pharmaceutical research. In this dissertation, we will focus on improving statistical designs of phase I and II trials to find solutions to several major obstacles in cancer clinical studies.

In cancer phase I trials researchers determine the dose toxicity relation by observing how the new drugs or drug combinations affect patients within a predefined time frame. A new patient will not be enrolled in the trial until all current observations obtain their information completely for majority of dose-escalation methods. In some other trials, new accrual is made with simply omitting the most recent incomplete observations; and assessment of the toxicity is based on all acquired information. However, both accrual methods raise some issues. The former accrual plan complicates accesses of the patients. The repeated accrual suspensions will also lead to longer trial duration and higher administrative cost. The reduced trial duration in the later accrual plan causes a loss of dose-toxicity information.

Moreover, in most of all phase I trials the toxicity response is represented by a binary indicator — dose limiting toxicity (DLT), which is defined as the toxic effects that are considered to be severe enough to prevent further dose escalation. However, using the binary indicator DLT may cause a loss in valuable information of the dose toxicity effect. For example, the actual grade 4 non-reversible renal toxicity and grade 3 reversible neutropenia as both classified as DLT, but the former one is much more severe than later one (Chen, et al., 2012). Similarly, for non-DLTs, grade 0, 1 and 2 toxicities are not equal. Because cancer phase I trials

are relatively small and designed with very limited number of patients, all toxicity information is very precious and should be fully used to maximize the trial efficacy.

In recent years, more and more concerns are raised towards the assumption of population homogeneity. Usually, in a cancer phase I trial, patients from the population are assumed to have the same MTD. However, this assumption is considered to be violated as recent knowledge obtained by the pharmacokinetics and the genetics of drug metabolism. For example, impaired renal function can result in reduced clearance of carboplatin and a dosing formulae based on renal function was developed (Newell, 1994). Taking patients individual characteristics into account is a new direction in cancer phase I trial designs.

Additionally, in the conventional phase II trials, researchers assess the effectiveness of the new drugs or drug combinations by investigating how and to what extent the tumor sizes change. Usually, tumor shrinkages more than 50% is considered to be an objective response (Julka, et al., 2008). The response rate is defined as the proportion of objective responses to all responses, and is utilized as the primary endpoint to evaluate effectiveness of the anti-tumor activity for the new drug or treatment. But concerns over the choice of response rate as the primary endpoint in cancer phase II trials have been raised (Karrison, et al., 2007). First, loss of information is the direct result of using the categorized variable tumor response. For example, a tumor size shrinks of 40% and a tumor size increases of 80% are both identified as a non-objective response, however, the treatment effects might be significantly distinct. Additionally, the high failure rate in subsequent phase III trials also dues to the discrepancy of endpoints between phase II and phase III trials.

In this dissertation, therefore, we will focus on the above concerns that have been well discussed lately and seek for solutions to conquer these obstacles. To improve phase I trial

efficiency, we will propose a hybrid design that considers four main competing interests: (1) preserve the safety of patients take part in the trial; (2) take into account of time factor allowing incomplete observations; (3) fully utilize all toxicity information; (4) incorporate patient's characteristics in the dose-toxicity model. For phase II trials, we proposed a novel two-stage design with double screening stages to improve the conventional phase II trial designs from the following aspects: (1) use continuous tumor size changes as endpoint in first screening stage; (2) utilize PFS as endpoint in second screening stage enhance subsequent phase III trial success rate; (3) double screening allows early termination which reduces the trial duration when no promising results present in stage I.

1.4 A Outline Structure

The rest part of this dissertation will follow the structure listed below.

In section 2.1 we will review two major methodologies — adaptive and sequential methods that are frequently used in clinical trials. Additionally, section 2.2 explores statistical methods adopted in phase I clinical trials with particular attention given to rule based designs and model based designs. We will also review some commonly used statistical methods in phase II clinical trials in section 2.3, where different types of phase II designs will also be introduced and discussed.

In chapter 3 we will describe the first part of our research that had been done in the program. We will propose a hybrid design for cancer phase I clinical trials. In section 3.1, we will discuss the challenges for current phase I designs and the motivation to develop a new design method. In section 3.2, we will provide a detailed description of our base model, dose escalation method with overdose control, time to event method and the novel toxicity scoring system to treat toxicity response as a quasi-continuous variable; as well as the theoretical

foundation for this design. In the last section of this chapter — section 3.3, we will provide a description of how we set up the simulation scheme. Finally, we will summarize and discuss our simulation results.

In chapter 4 we will discuss the second part of this dissertation. In section 4.1 we will propose a design base on EWOC-NETS-TITE to incorporate patients' characteristics. And in section 4.2, we will provide a description of the simulation plan to assess the performance of the model. In section 4.3 and section 4.4 are the simulation results and conclusion.

In chapter 5 we will describe the third part of research that had been done in the program. In the first section — section 5.1, we will propose a novel two-stage phase II design with double screening stages, and will provide an introduction to the research motivation. In section 5.2, we will provide a detailed description of the design scheme that how do we set up the simulation model. In the subsequent section — section 5.3, we will apply our proposed design to a simulated trial to construct a comparison with the most popular used design—Simon's two-stage design for phase II clinical trials. This simulation study examines the performance of our proposed phase II design by comparing success rate and model sensitivities to Simon's design.

Finally in chapter 6, we will conclude our proposed designs, summarize the research we have done for cancer clinical trial topic. In addition, we will explore possible directions and challenges for further research.

2 A REVIEW OF PHASE I AND PHASE II CLINICAL TRIALS

2.1 Adaptive and Sequential Methods in Clinical Trials

As we know, sample size in classical clinical trial designs is usually fixed and schedule is set up without using ongoing information during the trials. However, there is an increasing pressure on pharmaceutical companies and clinicians to re-examine the classical designs because of the fact that increasing spending in biomedical research didn't result in an increase of success rate (Chow, et al., 2008). Therefore, adaptive and sequential designs have been increasingly popular in clinical studies in recent years, which make the studies more efficient, more informative and more flexible.

Both adaptive and sequential designs allow changes of one or more specified aspects in the design or statistical procedure after initiation without impairing the validity and integrity of the trial (FDA, 2010). The goal of using such methods is not only to be more efficient to identify clinical benefits of the new drug under investigation, but also to improve the success rate of process of drug development. Statistical procedures that allowed changing can be the criteria to select and evaluate eligible patients, dose level of the testing drug or treatment, and schedule of the study. Other modifiable aspects include, but not limited to: study endpoints, measurement of clinical response, formulation of study objectives into statistical hypotheses, calculation of minimum sample size, participant randomization, study monitoring with interim/futility analysis, statistical data analysis plan, and so forth. Modifications are made to improve the performance of a trial with prompt use accumulated data from ongoing trial as well as upcoming related information from the literature (Chen, et al., 2012).

2.1.1 Adaptive Design

Adaptive design has been well discussed recently. An adaptive design is a clinical trial design that allows clinicians to modify some aspects of the study using accumulating data while it continues, without compromising the scientific method (Gallo, et al., 2006). This method is commonly adopted by adaptive randomization design, group sequential design, sample size re-estimation design, play-the-winner-and-drop-the-loser design, adaptive dose-finding design, adaptive treatment-switching design, hypothesis-adaptive design, and adaptive seamless phase II/III trial design to modify some aspects, such as sample size re-estimation, addition or removal of a study arm, treatment switch, and so on. All these designs can be classified into two categories — Bayesian and frequentist approaches. The frequentist approach performs the modification of trials while controlling for type I and type II errors. The Bayesian approach allows adaption according to the predicted probability. Recently, Bayesian approaches are increasingly popular in clinical trial designs.

As we discussed in chapter 1, we focus only on the early phases trial designs in this dissertation. An early phase I trial is conducted to determine MTD that dose levels are chosen adaptively based on the response to the most recent dose. A phase II trial, which is conducted to assess the new drug or treatment efficacy, is often done in two or more stages. It is also considered to be an adaptive method, because the decision in later stage is made based on the response in the previous stage (Cook, et al., 2010).

In phase I clinical trials an ongoing trial can be formulated by a function that logistic function is often used to model many different dose toxicity response relationships. Here, we use EWOC approach as an example, which is an adaptive design for dose finding based on a binary

indicator of DLT. The logistic dose-toxicity model is defined by the probability of toxicity response y_i:

$$\Pr(y_i = 1|x_i) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

where we assume that $\beta_1 > 0$ so that the probability of DLT is an increasing function of dose, and update is made when new response data become available (Chen, et al., 2012).

As we mentioned above, a phase II design with multi-stages is also considered to be adaptive. Early stop feature is desired to reduce the trial duration and cost when the efficacy or futility of the testing drug or treatment becomes obvious. In most phase II cancer clinical trials, primary endpoint in one of the multi-stages (usually two-stages) is tumor response rate (Kramar, et al., 1995). Then the hypothesis is usually stated as:

$$H_0: p \leq p_0 \text{ vs. } H_1: p \geq p_1$$

where true response rate is denoted by p, p_0 is the uninteresting response rate, and p_1 is chosen to be new drug or treatment's target response rate.

A predefined type I error rate is utilized to control the null hypothesis. If the null hypothesis is true, then stage II is preceded and a statistical power should be specified in alternative hypothesis (Lin, et al., 2004). In a clinical trial, p_0 is easily to be identified but in contrast, p_1 is not. In the early 1960's, when anticancer treatment is not as efficient as of today, Gehan (Gehan, 1960) proposed to use $p_0 = 0$, and it is widely used for decades. Then in 1989, Simon's two-stage design was proposed, which generalize p_0 and initiate the idea "optimal" and "minimax" design (Simon, 1989). There are many designs proposed later extend Simon's two-stage design, such as planned versus attained design (Green, et al., 1992), optimal three-stage design (Chen, 1997), and optimal two-stage design for single-arm trial (Shuster, 2002), and so

on. Additionally, some Bayesian designs have also been proposed recently, for example, Bayesian two-stage designs for phase II clinical trials (Tan, et al., 2002).

2.1.2 Sequential Design

Sequential designs in clinical trials have some common characteristics with adaptive design. They both allow modifications during the conduct of trial to design or statistical procedures according to the accumulated data. But it is required to have a choice of a primary patient response and of corresponding test statistics, a choice of stopping rule and a framework for the analysis to perform a sequential clinical trial. In other words, a sequential method mainly refers to the sequentially monitoring of stopping criteria for futility and efficacy (Todd, 1999).

When a clinical trial design is sequential, then infinite repetitions of the same sequential scheme must be contemplated. This raises the issue of lack in valid analysis that results in the design is not formally organized. But minor distinctions from the rule will only lead to negligible inaccuracies of analysis. Therefore, it is common to establish a safety monitoring mechanism (usually a committee) for major studies of life-threatening diseases. Recently, an unavoidable conflict in all clinical trials between the welfare of patients in the study and the scientific goals of the investigators attracts researchers' attention. This conflict can be minimized if the stopping rule is designed to cease the trial when it is predictable that the study has harmful impact on patients' safety or benefit. Therefore, a sequential design has the advantage of less likely to be in conflict with ethical requirements than a fixed-sample design. The sample size of a sequential study will be a random variable, with a distribution depending on the true treatment difference and the stopping rule used. Therefore, a sequential design has the benefit of reduction in sample size (Whitehead, 1997).

However, disadvantages also exist in a sequential design. The major drawback of a sequential design is that if the therapies under investigation are similar in their efficacies, the trial may require more patients than fixed-sample analysis (Thompson, 1980). It is also stated by researchers that a reliable statistical model need to be formulated before conducting the trial which is sometimes impossible.

To conquer the above stated issues, considerable novel statistical researches have been conducted in the development of adaptive and sequential approaches, especially for early phase's clinical trials — phase I and II trials. However, only limited numbers of the newly proposed methods have actually been applied to the daily practice of real clinical trials (Chen, et al., 2012). In section 2.2 and section 2.3, we will review significant and popularly used adaptive and sequential methods that have been applied to phase I and phase II clinical trials and have had a high impact on the field of clinical trials.

2.2 Statistical Methodology of Phase I Clinical Trials

It must be agreed by investigators participant in clinical trial studies that designing a clinical trial is when the researchers are struggling or are caught in the middle of two-front war. One front of designing a clinical trial is driven by the requirement that the research effort should be productive; the other is driven by statistical concerns.

There are two goals in conducting a phase I trial in cancer research: (1) determine an optimal dose (recommended for phase II trial), and (2) preserve the safety of the treatment for each individual patient as well as avoid too many cases to be treated at a conservative dose level. The two goals are proposed because unlike researches conducted in most therapeutic areas, a phase I trial is usually the first step to test the new drug or treatment in human beings (Eisenhauer, et al., 2000).

It is usually considered that a phase I trial is one of the most important steps in a drug development process after laboratory and animal studies that confirm a therapeutic agent has potential curative effect. Usually, the number of subjects in a phase I clinical trial is relatively small, fifteen to thirty, sometime sample size can varies in the range of twenty to eighty. Dose-toxicity response is used to model and determine the optimal dose which is set up by the widely accepted assumption that the therapeutic effect of a drug depends on its toxicity and increases monotonically with its dosage level (Tourneau, et al., 2009). Higher doses are associated with both severe toxicity and better therapeutic effect. Therefore, a balance is to be achieved between toxicity level and therapeutic benefit (Fanouriakis, et al., 2011). Seeking for the balance of toxicity level and therapeutic effect is equivalent to searching for a maximum dosage of the new drug under development. A patient should be treated with this maximum dose in the proceeding phase II and phase III trials, at which level the patient can tolerate its associated toxicities but should be with close monitoring.

Among all toxicities patients experience, some are so severe that to escalate the dose level. When the side effects developed are severe enough during the study, investigators should prevent for further dose escalation; and these toxicities are called dose limiting toxicity (DLT). National Cancer Institute (NCI, 2010) published their fourth version of Common Toxicity Criteria; DLT is defined as a group of grade 3 or higher non-hematologic toxicities and grade 4 hematologic non-transient toxicities. Table 2.1 lists the grade levels of all toxicities classified based on clinical descriptions of severity.

Designs for phase I cancer clinical trials have changed little in past decades. As mentioned above, the assumption is made that the higher the dose, the greater the likelihood of

efficacy. Therefore, in practical clinical studies dose-related toxicity is considered to be a surrogate for measurement of efficacy. Hence, the logic becomes

Table 2.1 Grade level for toxicity defined by CTCAE

Grade level	Toxicity	Clinical Description of Severity
Grade 0	No toxicity.	—
Grade 1	Mild toxicity.	Asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
Grade 2	Moderate toxicity.	Minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental activities of daily living.
Grade 3	Severe toxicity.	Server or medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care activities of daily living.
Grade 4	Life-threatening toxicity.	Life-threatening consequences; urgent intervention indicated.
Grade 5	Death.	Death related to adverse event.

the highest safe dose represents the most efficacious treatment effect. To estimate the maximum tolerated dose (MTD) of a new drug, a specific tolerable toxicity level needed to predefine. The highest acceptable DLT level is usually defined as a target toxicity level (TTL) (Parulekar, et al., 2004). The choice of the TTL is based on the preliminary knowledge of the new drug or treatment under investigation. If a new drug tends to have more server toxicity, then TTL selection should be more conservative. In other word, the choice of TTL determines the MTD of the new drug. Therefore, a careful and thoughtful approach to design phase I clinical trials and determine more accurate MTD is critical for the fate of the new drug in subsequent trials (Chen, et al., 2012).

For a cancer phase I trial, the guiding principle for dose escalation is to treat as many patients as possible at the therapeutic dose level to avoid unnecessary exposure of patients at the low doses of the new drug as well as preserve the safety. Thus, when prepare a dose escalation

scheme, we use the assumption that the probability of toxicity increases monotonically with increasing drug dose to establish the dose-toxicity relationship. Here in Table 2.2 we listed definitions of basic concepts used in phase I trials models from Tourneau e*t al.*, (Tourneau, et al., 2009), which will help us better understand the terminologies used later in this dissertation.

Table 2.2 Terms in common phase I clinical trial designs

Term	Definition
Cohort	Group of patients treated at a dose level.
Starting dose	The dose chosen to treat the first cohort of patients in a phase I trial.
Does increment (decrement)	The percent increase (or decrease) between dose levels.
Dose-limiting toxicity (DLT)	Toxic effects that are presumably related to the drug that are considered unacceptable (because of their severity and/or irreversibility) and that limit further dose escalation. DLTs are defined before beginning the trial and are protocol specific. They are typically defined based on toxic effects seen in the first cycle and specified using a standardized grading criteria, for example, CTCAE.
Dose-toxicity curve	The dose-toxicity curve reflects the relationship between dose and probability of toxicity for an anticancer agent. A logistic function is commonly assumed to describe the dose-toxicity curve for cytotoxic agents and is characterized by a parameter, θ, which represents the slope of the dose-toxicity curve. Small values of θ indicate that the probability of toxicity increase very slowly with increasing dose levels, whereas large values of θ indicate a sharp increase in toxicity with increasing dose levels.
Target toxicity level	The maximum probability of DLT that is considered acceptable in the trial. The TTL in phase I trials is typically between 20% and 33%.
Maximum tolerated dose (MTD)	Phase I trials conducted in the US: the highest dose level at which $\leq$ 33% of patients experience DLT. Phase I trials that use model-based methods: the dose that produces the target toxicity level (TTL).
Optimal biological dose (OBD)	Dose associated with a prespecified most desirable effect on a biomarker among all doses studied.
Recommended phase II dose	Phase I trials with a toxicity endpoint MTD.
Pharmacokinetics	Pharmacologic effects of the body on the drug, such as the time course of drug absorption, distribution, metabolism, and excretion.
Pharmacodynamics	Pharmacologic effects of the drug on the body, such as nadir neutrophil or platelet count, non-hematologic toxicity, molecular correlates, imaging endpoints.
Therapeutic index	The dosage or range of dosages of a drug that is required to produce a given level of damage to critical normal tissues (toxicity) divided by the dosage or range of dosages that yields a defined level of antitumor effect (efficacy).

Although it is popular to use an increasing relationship to describe the probability of toxicity, sometimes a decrease in the probability of toxicity at high dose levels could happen in some special cases. Since this type of cases is rare in practice, we will only focus on the common scenario in this dissertation. Dose escalation approaches for phase I cancer clinical trials can be divided in to broad classes: one uses nonparametric manner and is called rule-based design, including the traditional 3+3 design and its variations (Storer, 1989); the other is in the parametric way and called model-based design including CRM, EWOC, and their extensions. When the nonparametric design is adopted, the only assumption used to describe the dose-toxicity relationship is that toxicity is non-decreasing with dose. The rule-based designs assign patients to dose levels according to pre-specified rules based on actual observations of target events from the clinical data.

Sometimes the MTD or recommended dose for the subsequent phase II trial is also determined by the predefined rules. In contrary, the parametric method uses a model that adapts a distribution with some parameters to formulate the toxicity-dose curve. We can explain the mechanism of a dose-toxicity response biologically, that is the human body has stabilization and self-salvage systems to defend toxic invasion. The body will protect the person him/herself from mild toxicity when a drug dose is at a low level below a certain threshold level, but the probability of toxicity increases at an accelerated speed once the stabilization and self-salvage systems have been overcome, and reaches rapidly the worst condition, death, and then levels off. Therefore a sigmoid-shaped distribution is an appropriate model to describe the relationship between toxicity probability and dose. Many statistical designs have been proposed base on above considerations for phase I clinical trials; the most commonly used are summarized and compared in Table 2.3 (Chen, et al., 2012).

Table 2.3 Commonly used rule-based designs and model-based designs

Design class	Design	Advantages	Disadvantages
Rule-based designs	Standard 3+3 design	Robust. Simple. Easy to implement and safe.	Many patients treated at lower subtherapeutic doses. Slow dose escalation. MTD is not a dose with any particular probability of DLT, but in the range from 20% to 25% DLT. Cannot estimate MTD with target probability of DLT < 20% OR 33%. Not all toxicity data of all patients are used to determine the MTD.
	Isotonic design (ID)	Only assumes a monotonically increasing relationship between dose and toxicity. Semiparametric. Can estimate MTD with different TTL (0-100%). Robust and easy to implement. Good for combination of multiple drugs and treatments.	The accuracy of MTD may not be as good as CRM or EWOC. The trial efficiency may not be as good as CRM or EWOC.
Model-based designs	Continual reassessment method (CRM)	Fit parametric model for dose-toxicity relationship. Adaptive optimal design. Accurate estimation of MTD. Improved trial efficiency. Allow flexible MTD with different TTL.	High risk of patients being treated with over toxic dosages. If the parametric model is not reliable, the result could be questionable. May fail to find MTD.
	Escalation with overdose control (EWOC)	Includes all advantages of CRM. Controls the probability of overdosing a patient to toxic doses. Further improves MTD accuracy and trial efficiency.	If the parametric model is not reliable, the result could be questionable. May fail to find MTD.

2.2.1 Rule Based Designs

All the rule-based designs are considered to follow a sequential approach. It is very practical to make the assumption that there exists a non-decreasing dose-toxicity relationship,

which is the only assumption for a rule-based design. Although considerable new designs have been proposed by researchers in the past 20 years, rule-based designs are still playing a very important part (more than half of the clinical studies conducted) in cancer clinical study because of its feature — easy to implement. The most frequently used designs is standard 3+3 design (traditional 3+3 design) (Korn, et al., 1994), and isotonic design (ID) which was proposed (Leung, et al., 2001) to adjust the toxicity estimates using isotonic regression. Other popularly used rule-based designs include: standard A+B design, up-and-down design (Storer, 1989), and accelerated titration design (Simon, et al., 1997), and so on. In this section, we will review some of the commonly utilized rule-based designs.

The standard 3+3 design is used in phase I protocol templates of the cancer therapy evaluation program (CTEP), which primary goal is to improve the quality of lives of cancer patients by sponsoring clinical trials that investigate new anticancer drug or treatment with a particular emphasis on translational research to elucidate molecular targets and mechanisms of drug effects (Chen, et al., 2012). The standard 3+3 design can be described as follow. Assume a predefined dose range with ordinal dose levels $d_1 < \cdots < d_k$ and the corresponding probabilities of toxicity are $\Pr(\text{toxicity}) = 0 \leq q_1 < \cdots < q_k \leq 1$. The design is named 3+3 indicates the cohort size in the clinical trial is 3. The first cohort of 3 patients begins the trial at the starting lowest dose. Then simple version of dose escalation (without de-escalation) follows the scheme in Figure 2-1 with the following descriptions: for the i^{th} dose level,

1. Evaluate all 3 patients at d_i, when 0 out of 3 patients experience toxic event, then escalate to d_{i+1}. Otherwise, if no more than 1 patients experience toxic events, then go to step 2, if more than 1 patients experience toxic event, then terminate the trial, and recommend the proceeding dose d_{i-1} as the MTD.

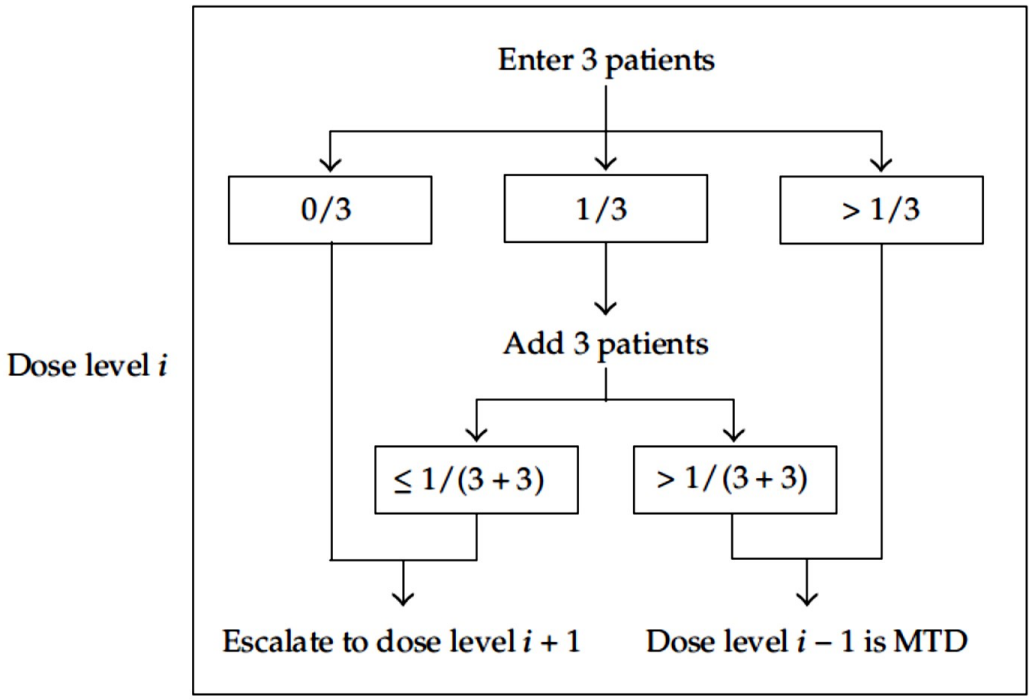

Figure 2-1 3+3 dose escalation scheme (Chen et al., 2012)

2. Evaluate an additional 3 patients at d_i, if no more than 1 toxic events are seen in total, then escalate; otherwise terminate the trial and recommend d_{i-1} as the MTD.

Although 3+3 design has been used as the protocol template in cancer phase I clinical trials, the MTD is not a dose with any particular probability of DLT. They are designed to screen drugs quickly and identify a dose level that does not exhibit too much toxicity in a very small group of patients. There are two types of 3+3 designs, the one we described above is 3+3 design without de-escalation, and there is another version with de-escalation that allows three new patients to be treated at a previous dose level. We will not discuss it here in this dissertation.

The isotonic design improves the standard designs in two ways. When summarizing toxicity risk at a dose, the method uses information at other doses in contrary to a standard design. And the design adopts isotonic regression which makes use of the monotonic character of

the dose-toxicity relationship. The design scheme adjusts 3+3 design is described by Leung *et al.*, as follow (Leung, et al., 2001), for dose level d_i,

1. Treat a cohort of 3 patients at d_i.
2. Evaluate toxicity at different doses, and choose the dose at which $\hat{q}_i$ is closest to the tolerable toxicity 0.33, where i is the dose last used. When $\hat{q}_i < 0.33$, then the dose escalates if smaller difference detected with next dose level d_{i+1} with $\hat{q}_{i+1}$ than with $\hat{q}_i$. That is $0.33 - \hat{q}_i \geq \hat{q}_{i+1} - 0.33$, otherwise continue at the same dose level. When $\hat{q}_i \geq 0.33$, then de-escalate the dose if $0.33 - \hat{q}_{i-1} \geq \hat{q}_i - 0.33$, otherwise continue at the same dose.

Many researchers have applied the isotonic designs to their clinical studies (Stylianou, et al., 2002) (Yuan, et al., 2004), the pool-adjacent-violators algorithm (PAVA) and isotonic regression are used in the method to update the probability of DLT of each dose level after obtaining the information that patients experience the toxic events. Because of its feature of model-free, it is suitable to cases where the parametric dose-toxicity relationship is not well understood (Chen, et al., 2012). However, simulation studies shows that using one-parameter model of isotonic design tends to overestimate the dose, and a very careful and thoughtful choice on the parameter should be made based on Markov chain theory to ensure more frequent assignments to the MTD and nearby doses (Ivanova, et al., 2009).

Rule-based designs, in practice, are still popular although many model-based designs have been proposed to cancer phase I trials. The simplicity of operating process is the reason to adopt these designs. There are many other rule-based designs have not been mentioned in this dissertation. All these rule-based designs can be used to determine a reasonable MTD using a stopping rule based either on observed DLTs or on convergence criteria. Ad hoc additional dose

levels can also be added when needed without any impact on their robustness. Rule-based designs thus are well suited for first in human clinical trials in which the dose-toxicity relationship is not well understood (Chen, et al., 2012).

2.2.2 Model Based Designs

In this section, we will provide an introduction to phase I model-based designs and a specific example will give to CRM (O'Quigley, et al., 1990) because in the later chapter 3, we will propose our hybrid design based on EWOC and provide a detailed description of it. As mentioned above, the body will defend the toxicity when initiating a drug dose then there is a steep increase in the probability of toxicity once the stabilization and self-salvage systems have been overcome. Thus a sigmoid-shape distribution is a very practical choice to build the dose-toxicity relationship. Among the model-based designs, three parametric dose-toxicity functions (logistic model, hyperbolic model, and power function) are frequently used to describe the relationship between dose and toxicity (Kang, et al., 2001).

A model-based design is a statistical model used to determine a dose level which could produce a prespecified probability of dose-limiting toxicity by using all enrolled patients in computing a more accurate dose-toxicity cure. The design is an alternative choice for the dose escalation method in cancer phase I trials. Recently, many multi-stage model-based designs have been proposed, for example two-stage design to optimize cancer phase I trials (Zandvliet, et al., 2010). But CRM (O'Quigley, et al., 1990) and EWOC (Babb, et al., 1998) are the ones that used most frequently. These two designs both use Bayesian approach to fully and efficiently utilize all available data and prior information in the phase I study. To use the CRM design, initial estimate of θ — TTL is required to set up before the clinical study. This will need inputs from experts who are familiar with the preclinical data or who have experience with similar drugs. Sometime,

the estimate may not be very accurate, but it provides a guidance for the dose escalation (Tourneau, et al., 2009).

We can describe the CRM design scheme as follows:

1. Establish a dose-toxicity model $\psi(x_i, a)$, where x_i the predefined dose is levels; a is a model parameter, and a prior distribution for parameter a (exponential distribution is often used) is assumed. Then define the target probability of θ (usually 20-33%).
2. At the dose level d_k, using the indicator y_k denotes the toxic event, and the posterior distribution of a is updated by Bayesian rule shown as follows:

$$f(a, \Omega_{k+1}) = \frac{g(a)\prod_{i=1}^{k}[\psi(x_i, a)]^{y_i}[1-\psi(x_i, a)]^{1-y_i}}{\int_0^{\infty} g(u)\prod_{i=1}^{k}[\psi(x_i, u)]^{y_i}[1-\psi(x_i, u)]^{1-y_i}du}$$

However, there is a controversial issue has been discussed in the CRM design. The CRM has a higher probability to overdose a patient comparing with the standard designs (Ratain, et al., 1993). Therefore, modified versions of CRM were proposed to avoid the overdosing problem. EWOC is the design that also can be considered as a translation from CRM. The advantage of this design is that it prevents patients from overdosing by controlling the probability of toxicity at a feasible bound. We will use the EWOC as a base model to implement our idea for the cancer phase I trial.

2.3 Statistical Methodology of Phase II Clinical Trials

There are large numbers of new cancer therapies emerge over recent years, but only a small number of the new drugs or treatments can be approved for market. Comparing with 20% success rate in cardiovascular new drug development, cancer drug attrition rates are significantly higher, only 5 out of 100 anticancer drugs under development would be approved (DiMasi, et al., 2007) (Walker, et al., 2009). These depressing statistics have led to increasing attention on the

drug development process, aiming to identify ways of reducing the failure rate. In reviewing the whole process, researchers found phase II trials act as screening tools before entering the most costly phase III trials (Brown, et al., 2011). Therefore, it is crucial for researchers to design the trail to give more accurate predict for the subsequent phase III trials. A phase II clinical trial is conducted to assess the new drug or treatment therapeutic effects using MTD recommended from proceeding phase I trial (Seymour, et al., 2010). The result provides critical information to determine whether conducting the large confirmatory phase III trials or not.

There are two types of phase II trials grouped by design method: single-arm designs and randomized designs; both can be easily understood from the names of category (Seymour, et al., 2010). In this chapter, we will review these two types of designs and construct a comparison of popularly used phase II designs which is described in Table 2.4 adopted from Chen *et al*., (Chen, et al., 2012). On the other hand, there is another way to specify a phase II trial by its ultimate aim, although the classification method is not quite popular. Divided by goals of the designs, phase II trials can be grouped as phase IIa (concept screening) and phase IIb studies (decision-making). Phase IIa trials aim to screen out the promising novel experimental agent with significant antitumor activity and phase IIb trials are conducted for seeking sufficient robustness to support progression to phase III studies.

The essential topic for designing a phase II trial is how to choose the most appropriate primary endpoint. Usually, response rate is chosen as the primary endpoint in a phase II trial. It was considered appropriate when unambiguous and clinically relevant antitumor activity is hypothesized. But recently, more and more surrogate endpoints have been proposed to evaluate the clinical benefit and treatment effects. Progression-free survival (PFS) of treated patients is one of the often used endpoints among all.

Table 2.4Common designs of phase II clinical trials

Design	Advantages	Disadvantages
Gehan's two-stage design	With interim monitoring. Rule out ineffective drug with minimized sample size.	No testing on agents showing some promise. Only suitable for binary outcome. The endpoint is different from that in following phase III trial.
Simon's two-stage design	The samples in two stages are optimized. Quickly screen out agents without effectiveness while testing further agents with some promise. Two choices: optimal vs. minimax.	Only suitable for binary outcome. The endpoint is different from that in the following phase III trials.
Randomized phase II design	Use of randomization. Reliable control and less bias. More similar to phase III trials.	Sample size increases. Length of trial increases. Cost increases
Phase II pick the winner design	Efficient and effective way of comparing two or multiple experimental regimens. Each experimental regimen compared with historical controls.	Not appropriate for comparison of adding an experimental agent to standard regimen.

2.3.1 Standard Single-Arm Designs

The standard single-arm designs are most commonly used in phase II clinical trials. Single-arm is utilized to compare the new drug or treatment with the standard response rate reported by historical data. As we mentioned before, the null hypothesis of such a design tests of insufficient efficacy while the alternative hypothesis test whether the new drug or treatment has sufficient activity to warrant further investigation. The type I error thus defines the chance that an ineffective agent will be studied further, and the type II error specifies the chance that an effective agent will not be studied further, which is usually regarded as the more serious error in phase II testing. In a study of mono-therapy, an agent might be considered uninteresting if the true response rate is $\leq$5% and interesting if the response rate is at least 20%; this criterion varies according to studies (Gray, et al., 2006).

A single-arm phase II trial design can have two or more stages, with the purpose to improve the trial efficiency and save resources by the early termination decision. Gehan first proposed a two-stage phase II design allowing early termination when no patients enrolled in stage I show any objective response. Otherwise, the trial continues to stage II which provides a more accurate response rate with additionally recruited patients (Gehan, 1960). However, there is no statistical testing on agents in Gehan's two-stage design and is not optimized. Simon (Simon, 1989) thus proposed an optimized two stage phase II design that controls both type I and type II errors as well as optimizes the sample sizes in both stages. The benchmark mentioned in previous paragraph is often tested with Simon's two-stage design. This design can quickly filter out ineffective drugs or treatments while further test more promising ones. The design has two subtypes, optimal and minimax. The optimal subtype minimizes the expected overall sample size with the probability to stop the trial after the first stage. It is appropriate to model experimental drugs efficacy which ones with a high probability to fail after the first stage (Chen, et al., 2012).

Researches interests have been attracted to how to determine appropriate endpoint and levels of activity for single-arm studies. PFS has been used as a surrogate endpoint in some recent phase II trials because of its advantages of short follow-up time and best estimate for overall survival (OS). Another commonly used surrogate endpoint is the continuous tumor change percentage. Traditionally, the categorical variable — tumor response is the most common endpoint in the phase II clinical trial designs. However, from a statistical point of view, categorizing a continuous tumor change percentage into a categorical tumor response with 4 levels results in a loss of statistical power by not fully utilizing all available data. We will focus on these well discussed topics in chapter 4 and propose a two-stage to overcome the above stated

problems, also compare the performance of our design with the standard protocol of most cancer phase II trials — Simon's two-stage design.

2.3.2 Randomized Phase II Design

There are several types of randomized phase II designs, include randomized control phase II design, randomized selection phase II design, and randomized discontinuation design, and so on. A randomize-controlled phase II study design typically compares an experimental regimen to a control arm (with or without a placebo) (Korn, et al., 2001). Comparison to a control arm is very useful when there is little prior information on expected efficacy rates in a population; and also when endpoints that are heavily influenced by patient selection, for example when PFS is used as an endpoint (Simon, et al., 2001). The endpoint could be a standard measure of tumor status, such as response rate or PFS which allows the study to be completed with fewer patients than required in a phase III study of survival. Type I and type II error rates are different as in other phase II designs. And the type I error can generally be larger than the one used in phase III studies; Korn *et al.* (Korn, et al., 2001) suggests considering one-sided type I error rates as large as 20%. The magnitude of the difference between the null and alternative hypotheses may also be larger than an appropriate value for a phase III study.

A randomized phase II selection design allows multiple single-arm studies to be conducted in the same time frame and with the same entry criterion (Lee, et al., 2005). The advantages of a randomized study over separate studies include decreasing the effects of patient selection bias, population drift and stage migration, and the ability to ensure that uniform evaluation criteria are used (Simon, et al., 1985). Although these studies are often designed to evaluate each arm separately and there is generally not adequate power for formal testing to compare arms, a predetermined plan for selection of arms for future study can still be made in

such a design (Scher, et al., 2002). Typically, this design randomizes between two or more experimental arms without a control arm. The advantage of the design is less selection bias due to changing natural history. A weakness of this design is the reduced likelihood of being able to select the best arm with increasing number of arms in the study or if there is a small difference in activity among arms (Gray, et al., 2006). There are many other designs, such as randomized discontinuation designs, factorial designs, and Bayesian designs, proposed by researchers for phase II clinical trials. Still, the main obstacle in front of current phase II trial designs is how to improve the follow up phase III trial success rate, which will be further discussed in chapter 4.

3 HYBRID PHASE I DESIGN — EWOC-NETS-TITE

The development of new anticancer drugs is a very complicated, extremely expensive and time-consuming process which includes discovery of agents that demonstrate antitumor activity in preclinical models and evaluation of normal tissue toxicity and application and confirmation on human beings in clinical trials (Ratain, et al., 1993). The primary goal of the initial cancer clinical trials step — phase I clinical trials is to determine the dose-toxicity relationship and recommend/schedule a maximum tolerate dose (MTD) for the subsequent phase II trials (Ho, et al., 2006). Most of the new drugs or treatment regimens that studied in phase I trials continue to phase II studies, which are conducted to evaluate the new drugs or treatments' antitumor activity. Drugs proceeded in phase I trials are rarely withdrawn at the completion of the testing (Joffe, et al., 2006).

In traditional cancer clinical studies, a phase I trial is the first step that a new drug or treatment under investigation applies to human beings. Since information about the bodies' reaction to the drug is very limited, the safety and ethical issues is in the highest priority. The purpose of conducting a phase I trial is to seek for new drug's toxic effect on patients as well as looking for an optimal dose that could balancing the maximum treatment effect and a tolerable toxic effect (Ishizuka, et al., 2001). Because the general assumption behind is the higher the dose, the better the treatment effect. But everything has its two sides; the opposite side of a higher dose is the probability of more severe toxic effect. Therefore, one of the well discussed and also the essential topic in cancer phase I trial designs is how to determine the dose-toxicity effect fast and accurately.

Patients who take part into a cancer phase I trial are usually the ones could not find promising effect on standard treatment. Rarely have patients volunteered to participate in the

phase I trials due to the treatment effects have not been proven. The sample size thus is small among all three phases, usually less than 30 patients. Therefore, another research area in phase I trial design is how to use such limited number of patients figure out the MTD effectively. This requires clinicians to quickly escalate the dose level to avoid assigning many patients to subtherapeutic dose levels as well as control too fast escalation which may result in overdosing patients.

The last but not the least important topic is how to reduce the trial duration. It seems less important than the above two topics. However, in practice, cancer patients especially patients participant in phase I trials with advanced cancer types cannot bear a long waiting in the trial without given any treatment intervention. There are two common methods to assign the new recruiting patients. One is directly assigning the new patient to a waiting list until completely observed all current patients, and then updating the patient dose level using all previous information. This method definitely has advantage at estimating a more accurate MTD, but sometimes the patients' waiting time is too long to hold them stay in the trial. The other commonly used method is simply omitting the incomplete observations, using only complete follow-ups to update the dose level. The advantage of this method is a reduced waiting time and trial duration, but may lead to an inaccurate MTD by repeating low conservative dose levels.

All above discussed topics motivate us searching for a method that could improve a phase I trial from these three areas: fast dose escalation, accurate MTD estimate, and short waiting time. Actually, these are also the universal goals for cancer phase I trial designs. We will apply the current optimal resources to it. But definitely, there is still long way to go in the further.

3.1 Introduction

In this section, we will review some current advanced designs in cancer phase I trials. And later, in the next section we will propose a hybrid design based on them. As mentioned previously, a phase I clinical trial is a key step in anticancer drug development, aiming at identifying the dose to be recommended for further phase II clinical trials. The primary purpose of cancer phase I clinical trial which is a critical step in development of new drug against cancer is to determine the maximum tolerated dose (MTD) and schedule of new drug. It is usually a small study with limited data so that fully utilizations of all toxicities and time to toxicity data are essential to improve the trial efficiency and accuracy of MTD estimation. A novel normalized the equivalent toxicity score (NETS) system has been proposed which can fully utilize multiple toxicities per patient instead of a binary indicator of dose limiting toxicity (DLT) (Chen, et al., 2010). The time of event (TITE) approach has developed to incorporate time to toxicity data (Cheung, et al., 2000). Escalation with Overdose Control (EWOC) is an adaptive Bayesian phase I design which allows rapid dose escalation as well as control the probability of overdosing patients (Babb, et al., 1998). In this study, we use EWOC as a framework and integrate it with the NETS system and TITE approach to develop an advanced phase I design entitled EWOC-NETS-TITE. This hybrid design can not only improve the trial efficiency and MTD accuracy substantially, but also allow patients to be entered in a staggered fashion and shorten trial length.

3.1.1 Dose Escalation with Overdose Control

Cancer phase I trials are carried out sequentially, assigning dose levels to subjects based on the observed side effects of the previously treated patients (Cheung, 2005). From a safety and therapeutic perspective, these trials should be designed to minimize the number of unacceptable

toxic events and maximize the number of patients treated at an optimal dose (Tighiouart, et al., 2006). Therefore, the design should control the probability of overdosing patients at each stage of the trial, produce a sequence of doses that converge to the MTD, and should take into account the heterogeneous nature of cancer phase I trial patients. There are several proposed designs for phase I trials that estimate MTD with a pre-specified probability, such as the continual reassessment method (O'Quigley, et al., 1990), the biased coin design (Durham, et al., 1997) and escalation with overdose control (Babb, et al., 1998). The basic ideas of these methods can be described in general as:

Let y_i be a binary random variable of toxicity observation (equal to 1 if DLT, 0 otherwise) and x_i be the dose level administered to the i^{th} patient, where $x_i \in \mathcal{D}$ and $\mathcal{D} = \{d_i : i = 1, \dots, k, d_1 < d_2 < d_k\}$ denote a pre-specified spaces dose levels that would be used in the escalation. Therefore, the pair (x_i, y_i) be the sequential dual dose levels and toxicity observations in the trial for i^{th} patient. The trial dose range is also pre-defined within $[x_{\min}, x_{\max}]$ with preliminary information obtained of the drug toxicity effect. Then the estimated MTD (γ) is defined as the dose associated with a specified proportion of patients experiencing a DLT, (θ)

$$\Pr(y_i = 1 | x_i = \gamma) = \theta$$

where θ depends on the context of the trial and the nature of the expected toxicities, usually the more severe toxicities are, the lower the θ is. The EWOC method is established on the basic idea mentioned above. The dose escalation starts after the first cohort of patients, if patients did not develop DLT at the first dose $x_{\min}$. Decision to escalate or de-escalate dose level is made after each cycle of therapy to a cohort of patients. The length of an observation cycle or time window is usually between 3 and 6 weeks (Roberts, et al., 2004) (Rogatko, et al., 2008). Each subsequent

dose to be assigned is determined via an estimate of the dose-toxicity relationship conditionally to the known information Ω. That is the dose assigned to the i^{th} patient, the posterior cumulative distribution function of MTD (π_k) will be

$$\pi_i(\gamma) = \Pr(\text{MTD} \leq \gamma | \Omega_i)$$

As discussed in previous chapters, the dose-toxicity relationship is often modeled with logistic model, which is also used in EWOC. The probability of patient to develop DLT is

$$\Pr(y_i | x_i, \beta_0, \beta_1) = F(\beta_0 + \beta_1 x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

where β_1 is a non-negative value according to the assumption toxicity is monotonically increasing with dose level. Hence the likelihood function can be derived as follow given the observed Ω_k

$$L(\beta_0, \beta_1 | \Omega_k) = \prod_{i=1}^{k} F(\beta_0 + \beta_1 x_i)^{y_i} [1 - F(\beta_0 + \beta_1 x_i)]^{1-y_i}$$

In section 3.2, we will improve this model by adding TITE and NETS features.

3.1.2 Time-to-Event Approach in Dose-Finding Studies

As we discussed above, traditional phase I trials require complete observation of each patient or a cohort of patients followed up by the entire observation time window. The need of a waiting list complicates access of the patients to the phase I trial. Moreover, repeated accrual suspensions impose very excessive administrative burdens and inconvenient long trial duration. Usually, one way to solve the timeliness problem is to replicate dose levels when patients are accrued before the acquisition of new complete data. But this method will allocate the same dose level for several patients without any dose escalation and updating estimate of MTD when the accrual rate is high. The newly recruited patients may also be accrued at dose level recommended on the basis of all complete available information ignoring data of the last patients

if not completed. In both scenarios, a high proportion of patients can be included at inefficient dose levels. Moreover, as the number of patients in a phase I trial is very limited, these may lead to a poor estimate of MTD.

Another solution to overcome the problem stated above is the time-to-event approach (Cheung, et al., 2000). This approach was first applied to the CRM design in which it uses a weight function in dose-escalation decision during the trial. We will adopt this idea to the EWOC design in this dissertation.

The weight is a function of the actual assessment time of the patient, is defined to 1 for a complete observation, for example, a patient with observation time completed or with occurrence of a DLT. It is assumed that the times until toxicity, as a proportion of the planned assessment time window, have a Beta distribution that can reflect the occurrence of early- or late-onset toxicities without correctly specifying the actual distribution of toxicity times (Braun, 2006). The weight function is as follow

$$w(u_i; T) = \begin{cases} w_i = \dfrac{u_i}{T} & \text{if } y_i = 0 \\ w_i = 1 & \text{if } y_i = 1 \end{cases}$$

where the u_i is the actual duration of assessment for i^{th} patient when a new patient enters the trial, and T is the predefined observation window. This weight function might appear to be an oversimplified choice but has been shown of adequate in many cases through simulation studies (Cheung, et al., 2000).

3.1.3 Novel Toxicity Scoring System

In most cancer phase I clinical trials, toxicity response is reduced to be a binary indicator as 1 for DLT and 0 for non DLT. As introduced in section 2.2, DLT is defined as a group of grade 3 or 4 non-hematologic and grade 4 hematologic toxicities or more severe toxicity (death)

(Rosenberger, et al., 2002) (Potter, 2006). However, using this binary indicator is hard to model the complicated scenarios when patients experience multiple toxicities and when there are correlations exist between different toxicities. Moreover, even though patients' toxic effects identified with the same DLT level may not be equally severe. For example, a grade 4 non-reversible renal toxicity is much more severe than a grade 3 reversible neutropenia, although both identified as DLT (Bekele, et al., 2004). Consider the way binary indicator classifies the toxic effects on patients; it may conduct to poor estimates of the toxicity reactions. As it is known to all, cancer phase I trials are small, so all the information obtained is very precious including the toxicities should be fully utilized (Yuan, et al., 2007).

The toxicity scoring system (NETS) is a solution to the dichotomized DLT indicator (Chen, et al., 2010). NETS is a comprehensive toxicity scoring system calculate an equivalent score measuring the composite severity of multiple toxicities experienced by each patient for phase I clinical trials. A logistic function to model the toxicity scores because the range of the values (range from 0 to 1) fits the gap between two consecutive adjusted toxicity grades.

Calculation of the NETS can be formulated as follow:

Let $G_{i,k,j}$ be the j^{th} toxicity grade by NCI toxicity criteria for the i^{th} patient who receives dose d_k , where $d_k \in \mathcal{D}$ and $\mathcal{D} = \{d_k : i = 1, \dots, K, d_1 < \cdots < d_k\}$ denotes a predefined space of dose levels for escalation. To avoid confounding of NCI toxicity criteria of grade 3 or 4 non-DLT and grade 3 or 4 DLT, $G_{i,k,j}$ can be adjusted to a new grade $G'_{i,k,j}$ranging from 0 to 6.

Table 3.1Adjusted toxicity grade

NCT toxicity grade	Grade 0	Grade 1	Grade 2	Grade 3 non-DLT	Grade 4 non-DLT	Grade 3 DLT	Grade 4 DLT
Adjusted grade	0	1	2	3	4	5	6

The maximum adjusted grade among all toxicities of the i^{th} patient given dose d_k is denoted by $G'_{i,k,\max} = \max G'_{i,k,j}$. An equivalent toxicity score is defined as

$$s_{i,k} = G'_{i,k,\max} - 1 + \frac{\exp\left(\alpha + \beta\left(\sum_{j=1}^{J} \frac{w_i G'_{i,k,j}}{G'_{i,k,\max}} - 1\right)\right)}{1 + \exp\left(\alpha + \beta\left(\sum_{j=1}^{J} \frac{w_i G'_{i,k,j}}{G'_{i,k,\max}} - 1\right)\right)}$$

While the normalized equivalent toxicity score goes one step further

$$s^*_{i,k} = \frac{1}{6}\left[G'_{i,k,\max} - 1 + \frac{\exp\left(\alpha + \beta\left(\sum_{j=1}^{J} \frac{w_i G'_{i,k,j}}{G'_{i,k,\max}} - 1\right)\right)}{1 + \exp\left(\alpha + \beta\left(\sum_{j=1}^{J} \frac{w_i G'_{i,k,j}}{G'_{i,k,\max}} - 1\right)\right)}\right]$$

where w_i ranges from 0 to 1 is a weight, which represents the correlation between j^{th} toxicity and all other toxicities the i^{th} patient experienced. The choice of the weight needs the input from experts who well understand the drug under development. α is the impact distinction between the worst and other toxicities that have imposed on the patient, and α usually assumed to -2. β is nonnegative reflecting the increment in ETS by additional toxicity. Since a patient could only be given a single dose level during phase I clinical trials, so we can simplify the notation $s^*_{i,k}$ to s_i. Now, the NETS s_i is a value between 0 and 1. A summary of toxicity scoring system (Table 4.2) has been provided by Chen *et al.* shows the range of the NETS (Chen, et al., 2010).

Table 3.2 Summary of toxicity score system

Most severe toxicity NCI toxicity grade	Maximum adjusted grade $G'_{i,\max}$	Range of NETS s_i	Mid-range NETS
Grade 0	0	0	0
Grade 1	1	[1/60-1/6)	0.092
Grade 2	2	[1/6-1/3)	0.25
Grade 3 non-DLT	3	[1/3-1/2)	0.417
Grade 4 non-DLT	4	[1/2-1/3)	0.583
Grade 3 DLT	5	[2/3-5/6)	0.75
Grade 4 DLT	6	[5/6-1)	0.917

As discussed previously, target toxicity level — TTL (θ) is a prespecified value represents the probability of a patient that experiences DLT when given the MTD (γ). As DLT is not available, a new variable target normalized equivalent toxicity score — TNETS is used. Similar as TTL, the TNETS is also based on researchers' prior understanding about the new drug or treatment.

3.2 A Hybrid Design for Phase I Clinical Trials — EWOC-NETS-TITE

In this section, we propose a hybrid phase I design dose escalation method with overdose control using a normalized equivalent toxicity score system and time-to-event approach. The aim of design is to decrease the dose-finding trial duration, without impairing the characteristics of the EWOC design, especially the overdose control ability, as well as fully utilize all toxicity information to estimate MTD more accurately.

3.2.1 EWOC

The proposed design adopts EWOC method to model the dose escalation scheme. We will start from the EWOC method and still use DLT here to demonstrate the process, then apply NETS and TITE step by step. Suppose y_i is the indicator of DLT for the i^{th} patient given dose level at x_i. The probability of DLT is usually formulated with logistic model,

$$\Pr(y_i|x_i,\beta_0,\beta_1) = F(\beta_0+\beta_1 x_i) = \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}$$

We can re-parameterize this relationship in terms of MTD (γ) and probability of DLT at initial dose (ρ_0) to get this equation more practice meaning. That is,

$$\theta = F(\beta_0+\beta_1\gamma) = \frac{\exp(\beta_0+\beta_1\gamma)}{1+\exp(\beta_0+\beta_1\gamma)}$$

Hence, the estimated MTD γ and ρ_0 can be derived as

$$\gamma = \frac{F^{-1}(\theta) - \beta_0}{\beta_1} = \frac{\ln(\theta) - \ln(1-\theta) - \beta_0}{\beta_1}$$

$$\rho_0 = F(\beta_0 + \beta_1\gamma) = \frac{\exp(\beta_0 + \beta_1 x_{\min})}{1 + \exp(\beta_0 + \beta_1 x_{\min})}$$

Then the previously used intercept β_0 and slope β_1 can be written as follow:

$$\beta_0 = \frac{x_{\min} \cdot \text{logit}(\theta) - \gamma \cdot \text{logit}(\rho_0)}{x_{\min} - \gamma}$$

$$\beta_1 = \frac{\text{logit}(\rho_0) - \text{logit}(\theta)}{x_{\min} - \gamma}$$

The second equation shows that the assumption of non-decreasing dose-toxicity relationship that $\beta_1 > 0$ implying $0 < \rho_0 < \theta$. Thus, The likelihood function can be updated by using the two parameters γ and ρ_0, which is

$$L(\rho_0, \gamma | \Omega_j) = \prod_{i=1}^{j} F(\rho_0, \gamma, x_i)^{y_i} [1 - F(\rho_0, \gamma, x_i)]^{1-y_i}$$

Expand the function $F(\rho_0, \gamma, x_i)$ that is,

$$F(\rho_0, \gamma, x_i) = F\left(\frac{x_{\min} \cdot \text{logit}(\theta) - \gamma \cdot \text{logit}(\rho_0)}{x_{\min} - \gamma} + \frac{\text{logit}(\rho_0) - \text{logit}(\theta)}{x_{\min} - \gamma} x_i\right)$$

The likelihood function can be written as

$$L(\rho_0, \gamma | \Omega_j) = \prod_{i=1}^{j} \left(\exp\left\{\frac{(\gamma - x_i) \cdot \text{logit}(\rho_0) + (x_i - x_{\min}) \cdot \text{logit}(\theta)}{\gamma - x_{\min}}\right\}\right)^{y_i}$$

$$\times \left(1 + \exp\left\{\frac{(\gamma - x_i) \cdot \text{logit}(\rho_0) + (x_i - x_{\min}) \cdot \text{logit}(\theta)}{\gamma - x_{\min}}\right\}\right)^{-1}$$

After specifying the prior distribution $h(\rho_0, \gamma)$ for the pair of parameters (ρ_0, γ), we denote $\pi_n(\gamma)$ as the marginal posterior cumulative density function of γ given D_n, then the

EWOC method can be described as follow: The first patient receives the dose $x_1 = x_{\min}$ and conditional on the event $\{y_1 = 0\}$, the $(n+1)^{th}$ patient receives the dose

$$x_{n+1} = \pi_n^{-1}(\alpha)$$

so that the posterior probability of exceeding the MTD is equal to the feasibility bound α. While $y_1 = 1$, the trial should terminate in purpose of safety.

3.2.2 EWOC-TITE

Now, we apply the time-to-event approach in the EWOC design. The proposed TITE method introduces a weight w first in the CRM design (Cheung, et al., 2000) based on a logistic model taking into account time factor in terms of time until toxicity. With TITE, a weighted dose-toxicity relationship is denoted by $G(x_i, w_i, \beta_i)$ is monotone increasing in w with marginal constraints $G(x_i, 0, \beta_i) = 0$ and $G(x_i, 1, \beta_i) = F(x_i, \beta_i)$ for all x, β. The weight w is linearly defined, and impose into F, $G(x_i, w_i, \beta_i) = wF(x_i, \beta_i)$, where $0 \leq w_i \leq 1$.

The weight function can be selected from various relations; however, this relation should be chosen accommodate to the toxicity profile with respect to the drug under investigation and in accordance with the planned assessment time window. Considering the nature of a cancer phase I trial, when a patient experiences DLT, the observation is identified as a complete observation. The assessment time ends at the time of the emergence of DLT and the weight for this observation equals 1.

Incorporating the weight function into the dose escalation study allows us to take into account all available information, including incomplete information in the process of estimating dose-toxicity relationship. Therefore, a new patient is recruited when evaluation of the last patients is not completed, the patient can still enter the trial rather than stay in the waiting list; the allocated dose is estimated using all complete and incomplete information. With this method,

waiting lists can be avoided and all eligible patients can benefit from the clinical trial. And the likelihood function incorporates the TITE shown as follow:

$$L(w, \rho_0, \gamma | \Omega_j) = \prod_{i=1}^{j} [w_i \cdot F(\rho_0, \gamma, x_i)]^{y_i} [1 - w_i \cdot F(\rho_0, \gamma, x_i)]^{1-y_i}$$

which can be further expanded as

$$L(w, \rho_0, \gamma | \Omega_j) = \prod_{i=1}^{j} w_i^{y_i} \left(\exp\left\{ \frac{(\gamma - x_i) \cdot \text{logit}(\rho_0) + (x_i - x_{\min}) \cdot \text{logit}(\theta)}{\gamma - x_{\min}} \right\} \right)^{y_i}$$

$$\times \left(1 + (1 - w_i) \exp\left\{ \frac{(\gamma - x_i) \cdot \text{logit}(\rho_0) + (x_i - x_{\min}) \cdot \text{logit}(\theta)}{\gamma - x_{\min}} \right\} \right)^{1-y_i}$$

$$\times \left(1 + \exp\left\{ \frac{(\gamma - x_i) \cdot \text{logit}(\rho_0) + (x_i - x_{\min}) \cdot \text{logit}(\theta)}{\gamma - x_{\min}} \right\} \right)^{-1}$$

3.2.3 EWOC-NETS-TITE

Before adopting NETS into EWOC-TITE method, we need to construct the target normalized equivalent toxicity score (TNETS) at first. As mentioned above, determination of TNETS depends on the target toxicity profile which relies heavily on clinician's input (Chen, et al., 2010). Four related questions should be specified in order to define the target toxicity profile. It consists of proportion of patients who experience DLT when treated at the MTD and the target probability that adjusted grade l toxicity is the expected worst toxicity when a patient is given MTD. Then the target normalized equivalent toxicity score is defined as

$$\tilde{\theta} = \sum_{l=0}^{6} m_l \cdot p_l$$

where m_l is the mid-range of NETS (can be find in Table 3,2) and p_l is the target probability according to the maximum adjusted grade l toxicity. Then the MTD γ is defined as a dose corresponding to a prespecified target normalized equivalent toxicity score (TNETS) $\tilde{\theta}$

$$\text{TNETS}=\text{NETS}|_{\text{Dose}=\gamma} = \tilde{\theta}$$

For the dose-toxicity relationship, a logistic function is also appropriate (Figure 3-1).

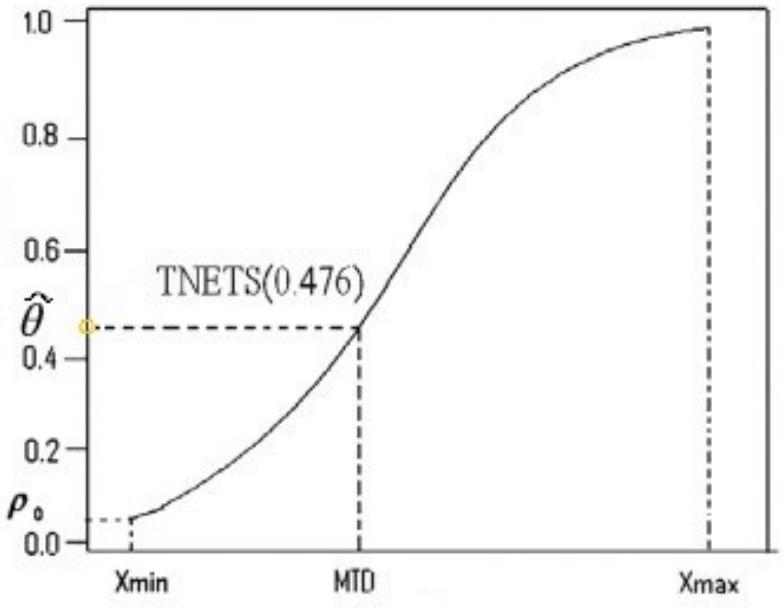

Figure 3-1Dose-toxicity relationship

Thus, the dose-toxicity model can be described as:

$$s_i = F(\beta_0 + \beta_1 x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1)}$$

where $\beta_1 > 0$ implies the toxicity score NETS is monotonically increasing with toxicity. We can also re-parameterize the model with respect to MTD γ and ρ_0, so that the interpretation of the model is explicit.

$$\tilde{\theta} = F(\beta_0 + \beta_1 \gamma) = \frac{\exp(\beta_0 + \beta_1 \gamma)}{1 + \exp(\beta_0 + \beta_1 \gamma)}$$

Thus, the estimated MTD γ and ρ_0 can be derived as

$$\gamma = \frac{F^{-1}(\tilde{\theta}) - \beta_0}{\beta_1} = \frac{\ln(\tilde{\theta}) - \ln(1-\tilde{\theta}) - \beta_0}{\beta_1}$$

$$\rho_0 = F(\beta_0 + \beta_1\gamma) = \frac{\exp(\beta_0 + \beta_1 x_{\min})}{1 + \exp(\beta_0 + \beta_1 x_{\min})}$$

The two parameters β_0 and β_1 can be written as follow:

$$\beta_0 = \frac{x_{\min} \cdot \text{logit}(\tilde{\theta}) - \gamma \cdot \text{logit}(\rho_0)}{x_{\min} - \gamma}$$

$$\beta_1 = \frac{\text{logit}(\rho_0) - \text{logit}(\tilde{\theta})}{x_{\min} - \gamma}$$

It can be shown that

$$\text{logit}\,(s_i|x_i) = \frac{(\gamma - x_i) \cdot \text{logit}(\rho_0) - (x_i - x_{\min}) \cdot \text{logit}(\tilde{\theta})}{\gamma - x_{\min}}$$

Thus, the likelihood function can be updated by using the two parameters γ and ρ_0, which is

$$\begin{aligned} & L(\rho_0, \gamma | D_j) \\ &= \prod_{i=1}^{j} c_i \cdot [w_i \cdot F(\rho_0, \gamma, x_i)]^{s_i} [1 - w_i \cdot F(\rho_0, \gamma, x_i)]^{1-s_i} \\ &= \prod_{i=1}^{j} c_i \cdot g(s_i) \end{aligned}$$

where $\int_0^1 c_i \cdot g(s_i) ds_i = 1$. With expansions and transformations, the likelihood function can be written as

$$L(\rho_0, \gamma | D_j)$$

$$= \prod_{i=1}^{j} w_i^{s_i} \cdot \frac{\ln\left[\left\{1 - w_i \exp\left(\frac{\text{logit}(\rho_o) \cdot (\gamma - x_i) + \text{logit}(\tilde{\theta}) \cdot (x_i - x_{\min})}{\gamma - x_{\min}}\right)\right\}^{-1} - 1\right]}{(2w_i - 1) \cdot \exp\left(\frac{\text{logit}(\rho_o) \cdot (\gamma - x_i) + \text{logit}(\tilde{\theta}) \cdot (x_i - x_{\min})}{\gamma - x_{\min}}\right) - 1}$$

$$\times \left[\exp\left(\frac{\text{logit}(\rho_o)\cdot(\gamma - x_i) + \text{logit}(\tilde{\theta})\cdot(x_i - x_{\min})}{\gamma - x_{\min}}\right)\right]^{s_i}$$

$$\times \left[1 + (1 - w_i)\cdot\exp\left(\frac{\text{logit}(\rho_o)\cdot(\gamma - x_i) + \text{logit}(\tilde{\theta})\cdot(x_i - x_{\min})}{\gamma - x_{\min}}\right)\right]^{1-s_i}$$

Using $h(\rho_0, \gamma)$ denote the prior distribution on $[0, \tilde{\theta}] \times [x_{\min}, x_{\max}]$, then the posterior distribution of (ρ_0, γ) is

$$\pi(\rho_0, \gamma | D_j) = \frac{L(\rho_0, \gamma | D_j)\cdot h(\rho_0, \gamma)}{\iint_{[0,\tilde{\theta}]\times[x_{\min}, x_{\max}]} L(\rho_0, \gamma | D_j) h(\rho_0, \gamma) d\rho_0 d\gamma}$$

Thus, $\pi_j(\gamma)$ is the marginal posterior c.d.f. of γ given D_j. We will conduct a simulation study using MCMC to update posterior distribution. The Metropolis–Hastings algorithm is implemented.

3.3 Simulation Study

A simulation study is conducted to evaluate the performance of our proposed hybrid phase I design EWOC-NETS-TITE. Since our design adopts the popularly used phase I trial design EWOC as a basic model, the result will mainly be used to compare with EWOC method. There are two major parts in the simulation study. First part of our simulation study is used to compare the accuracy in finding MTD, as the primary goal of a phase I clinical trial is to find an optimal dose. The comparison will be made with a series of designs include EWOC with complete observation of follow up, EWOC omit incomplete observation, EWOC-NETS, and EWOC-TITE. The second part of our simulation study is used to compare the trial durations with the above mentioned series of designs and evaluate whether the proposed design could effective reduce the trial duration as well as maintain a relatively accurate estimate of MTD. A replicate of 5,000 trials was done for both part of the simulation.

3.3.1 Simulation Plan

For the first part of our simulation, we conducted a simulation study to compare the model performance with a series of designs derived from EWOC method. The motivation to propose such a hybrid phase I design is to seek for a method that could effectively improve the a phase I clinical trial from two aspects: improve the MTD estimation accuracy by fully utilize the toxicity information, as well as reduce the trial duration trials avoiding long waiting list or simply dropping incomplete observations. With these purposes, we have done the following simulation:

(1) Comparison of accuracy in MTD estimation. Before the start of simulation study, specifying the target toxicity scenarios and predetermine a true MTD, so that the estimation could be compared with the true value of MTD. To set up a predefined MTD, there are several steps to complete. First, the TTL — probability of DLT of the estimated MTD needs to be determined. For example, a TTL is predefined as 33% in the design when treat toxicity response as a binary variable. Then the corresponding highest acceptable probability of DLT when treat toxicity response as a continuous variable, is also defined as 33%. We assume the probabilities of experiencing a grade 3 and grade 4 DLT are equal — chances are 1:1; and the probabilities of experiencing non-DLT grade 1 through grade 4 are also equal (1:1:1:1). The probability of experiencing a nontoxic effect would be set up lower than other probabilities according to reality. Therefore, corresponding to the 33% TTL, the target toxicity profile for each adjusted toxicity grade can be calculated. Furthermore, using both mid-range of the NETS and the target toxicity profile. The TNETS could be derived. Table 3.3 demonstrates the result of the scenario that is described above. The largest TNETS thus corresponds to the true MTD when trace back to the adjusted toxicity grade.

Table 3.3 TNETS calculation with sample target toxicity profile

Toxicity Grade	TTL Allocation	Ratio of Probability Allocation	Target Toxicity Profile	Mid-Range NETS	Contribution $m_l \cdot p_l$
Grade 0	67%	-	7.0%	0.000	0.0000
Grade 1		1	15.0%	0.092	0.0138
Grade 2		1	15.0%	0.250	0.0375
Grade 3 non-DLT		1	15.0%	0.417	0.0626
Grade 4 non-DLT		1	15.0%	0.583	0.0875
Grade 3 DLT	33%	1	16.5%	0.750	0.1238
Grade 4 DLT		1	16.5%	0.917	0.1513

Then the TNETS is $\tilde{\theta} = \sum_{l=0}^{6} m_l \cdot p_l = 0.476$. After obtaining the TNETS, we can get the true MTD by going through all average NETS (ANETS) at each dose level. The ANETS is calculated by

$$s_k^* = \frac{1}{n_k} \sum_{i=1}^{n_k} s_{i,k}$$

For different scenario setups, the ANETS varies. Then the MTD is the adjusted dose level with a minimum difference from the TNETS and ANETS.

In the simulation study, we used small sample sizes 1) 30 patients — a common phase I trial sample size, and 2) 60 patients to get more stable estimate of MTD; because large sample properties are not appropriate for assessing a phase I trial design. In addition, we compared different scenarios of target toxicity profiles. The simulation study was conducted to compare a series of EWOC derived designs with respect to the accuracy of MTD estimation.

Another aspect needs to be compared with is the trial duration. As discussed above, it is not practical to hold a cancer patient on a waiting list for long, especially for patients participant in phase I trials with advanced cancers. Moreover, it will burden the administrative system of the trials. So the goal of the simulation study is to seek for evidence that the hybrid design could effective reduce the trial duration.

3.3.2 Simulation Setup

Five different scenarios of target toxicity profiles are considered, with the probability of DLT being 33% that is commonly used in phase I trials. The target toxicity profiles can be found in Table 3.4-3.8 with different true MTD assumptions. These scenarios were adopted from Chen *et al*. The first scenario is considered to be "ideal", with an equal ratio for all non-DLT toxicities, and both DLT toxicities; and is called target toxicity scenario. With skewness to the lower toxicity grade, the second scenario can be described as under-toxicity scenario; in contrary the third over-toxicity scenario means the toxicity profile skews to higher toxicity grade with larger ANETS values. The true MTD and ANETS are highlighted in the tables.

Table 3.4 Target scenario for toxicity profile

Max. Adj. Grade	Probability the allocate a dose specific maximum adjusted grade						Mid-Range NETS
	1	2	**3**	4	5	6	
0	0.110	0.090	0.070	0.050	0.030	0.010	0.000
1	0.200	0.160	0.150	0.120	0.100	0.050	0.092
2	0.200	0.170	0.150	0.130	0.100	0.060	0.250
3	0.200	0.170	0.150	0.130	0.100	0.060	0.417
4	0.210	0.170	0.150	0.130	0.110	0.060	0.583
5	0.040	0.120	0.165	0.220	0.280	0.380	0.750
6	0.040	0.120	0.165	0.220	0.280	0.380	0.917
ANETS	0.341	0.427	**0.476**	0.540	0.607	0.713	

Table 3.5 Under-toxicity scenario for toxicity profile

Max. Adj. Grade	Probability the allocate a dose specific maximum adjusted grade						Mid-Range NETS
	1	2	3	**4**	5	6	
0	0.110	0.090	0.070	0.050	0.030	0.010	0.000
1	0.324	0.268	0.240	0.204	0.164	0.092	0.092
2	0.243	0.201	0.180	0.153	0.123	0.069	0.250
3	0.162	0.134	0.120	0.102	0.082	0.046	0.417
4	0.081	0.067	0.060	0.051	0.041	0.023	0.583
5	0.060	0.160	0.220	0.300	0.370	0.510	0.750
6	0.020	0.080	0.110	0.140	0.190	0.250	0.917
ANETS	0.269	0.636	0.410	**0.483**	0.556	0.670	

Table 3.6 Over-toxicity scenario for toxicity profile

Max. Adj. Grade	Probability the allocate a dose specific maximum adjusted grade						Mid-Range NETS
	1	**2**	3	4	5	6	
0	0.110	0.090	0.070	0.050	0.030	0.010	0.000
1	0.081	0.067	0.060	0.051	0.041	0.023	0.092
2	0.162	0.134	0.120	0.102	0.082	0.046	0.250
3	0.243	0.201	0.180	0.153	0.123	0.069	0.417
4	0.324	0.268	0.240	0.204	0.164	0.092	0.583
5	0.020	0.080	0.110	0.140	0.190	0.250	0.750
6	0.060	0.160	0.220	0.300	0.370	0.510	0.917
ANETS	0.408	**0.486**	0.526	0.593	0.653	0.751	

Table 3.7 Over-toxicity scenario of 33% TTL and grade 3 true MTD

Max. Adj. Grade	Probability the allocate a dose specific maximum adjusted grade						Mid-Range NETS
	1	2	**3**	4	5	6	
0	0.110	0.090	0.070	0.050	0.030	0.010	0.000
1	0.324	0.268	0.240	0.204	0.164	0.092	0.092
2	0.243	0.201	0.180	0.153	0.123	0.069	0.250
3	0.162	0.134	0.120	0.102	0.082	0.046	0.417
4	0.081	0.067	0.060	0.051	0.041	0.023	0.583
5	0.060	0.160	0.220	0.300	0.370	0.510	0.750
6	0.020	0.080	0.110	0.140	0.190	0.250	0.917
Prob. of DLT	0.080	0.240	**0.330**	0.440	0.560	0.760	
ANETS	0.110	0.090	**0.410**	0.483	0.556	0.670	

Table 3.8 Under-toxicity scenario for 33% TTL and grade 3 true MTD

Max. Adj. Grade	Probability the allocate a dose specific maximum adjusted grade						Mid-Range NETS
	1	2	**3**	4	5	6	
0	0.00	0.00	0.00	0.00	0.000	0.000	0.000
1	0.00	0.00	0.00	0.00	0.000	0.000	0.092
2	0.00	0.00	0.00	0.00	0.000	0.000	0.250
3	0.00	0.00	0.00	0.00	0.000	0.000	0.417
4	0.92	0.76	0.67	0.56	0.446	0.240	0.583
5	0.00	0.00	0.00	0.00	0.000	0.000	0.750
6	0.08	0.24	0.33	0.44	0.554	0.760	0.917
Prob. of DLT	0.08	0.24	**0.33**	0.44	0.560	0.760	
ANETS	0.61	0.66	**0.69**	0.73	0.770	0.840	

We use the MCMC sampling method to simulate the Bayesian framework of the two parameters (ρ_0, γ) of interest. The assessment time window was set up to 28 days, which is set up according to common clinical trials observation time (usually four weeks or one month). Time and events were simulated using an exponential distribution. We investigated different means of patient arrival time: 7, 28, and 100 days.

3.4 Simulation Results

In the first part of this section, we will demonstrate the performance of our two-stage design by comparing a series of EWOC derived methods that include EWOC-NETS (wait, cohort size 1), EWOC-NETS (wait, cohort size 3), EWOC-NETS (no wait), TITE-EWOC, EWOC (wait, cohort size 1) and EWOC (wait, cohort size 3). The later part of this section is the simulated trial duration comparing to these designs by different sample sizes.

3.4.1 Estimation of MTD

The purpose of conducting a phase I trial is to find the MTD and recommend for further phase II trial investigation. So the accuracy of the MTD estimation is an essential criterion for evaluation of the design performance. As we mentioned above, there were several EWOC derived methods used to construct the comparison. The reason to choose only EWOC derived methods, not CRM, not ID is: it has been well discussed that EWOC has its unique benefit superior to the other two popular used designs. Table 3.9 - Table 3.12 illustrates the recommended MTD in different scenarios for sample size 30 and 60, respectively. The toxicity profile is selected equals to 33% for target toxicity level (TTL), and then all target NETS (TNETS) are solved as 47.6%. Hence, for the three different scenarios, the MTD recommended is considered more accurate if the recommendation falls into the dose level with ANETS closes to 47.6%.

Table 3.9 Percent of MTD recommended according to ANETS = 47.6%

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	0	9	67	24	0	0
EWOC-NETS (wait, cohort size 1)	0	7	69	24	0	0
EWOC_NETS (wait, cohort size 3)	0	5	73	22	0	0
EWOC-NETS (no wait, cohort size 1)	0	5	64	31	0	0
EWOC-TITE	22	13	42	23	0	0
EWOC (wait, cohort size =1)	1	16	63	20	0	0
EWOC (wait, cohort size =3)	1	18	58	23	0	0
Under-Toxicity Scenario	Dose Level					
	1	2	3	**4**	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.27	0.36	0.41	**0.48**	0.56	0.67
EWOC-NETS-TITE	0	0	20	77	3	0
EWOC-NETS (wait, cohort size 1)	0	0	22	76	2	0
EWOC_NETS (wait, cohort size 3)	0	0	16	79	5	0
EWOC-NETS (no wait, cohort size 1)	0	0	24	69	7	0
EWOC-TITE	30	5	19	36	10	0
EWOC (wait, cohort size =1)	1	2	25	63	9	0
EWOC (wait, cohort size =3)	0	2	16	69	13	0
Over-Toxicity Scenario	Dose Level					
	1	**2**	3	4	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.41	**0.49**	0.53	0.59	0.65	0.75
EWOC-NETS-TITE	0	43	55	2	0	0
EWOC-NETS (wait, cohort size 1)	0	46	54	0	0	0
EWOC_NETS (wait, cohort size 3)	0	47	51	2	0	0
EWOC-NETS (no wait, cohort size 1)	0	41	56	3	0	0
EWOC-TITE	28	45	26	1	0	0
EWOC (wait, cohort size =1)	3	56	38	3	0	0
EWOC (wait, cohort size =3)	5	60	33	2	0	0

As we discussed above, the guiding principle for phase I clinical trials is to treat as many patients as possible to the maximum tolerate dose level to avoid giving patients the subtherapeutic dose (lower) levels while preserve safety. Hence, another important standard to assess a phase I design is how patients were allocated to the dose levels. The more patients were

treated at the optimal doses, the more efficient the trial is. Table 3.10 illustrates the how patients distribute for each dose level.

Table 3.10 Percent of patients treated at each dose level

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	2.82	17.25	52.53	26.57	0.83	0.00
EWOC-NETS (wait, cohort size 1)	4.27	18.07	47.58	28.08	2.00	0.00
EWOC_NETS (wait, cohort size 3)	5.15	16.80	50.85	26.10	1.10	0.00
EWOC-NETS (no wait, cohort size 1)	3.27	10.92	60.72	24.97	0.13	0.00
EWOC-TITE	31.82	12.67	26.97	24.03	4.52	0.00
EWOC (wait, cohort size =1)	6.80	25.37	44.33	21.37	2.13	0.00
EWOC (wait, cohort size =3)	8.75	28.35	37.50	22.05	3.35	0.00
Under-Toxicity Scenario	Dose Level					
	1	2	3	**4**	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.27	0.36	0.41	**0.48**	0.56	0.67
EWOC-NETS-TITE	1.68	5.15	34.45	53.35	5.37	0.00
EWOC-NETS (wait, cohort size 1)	1.67	3.53	33.90	55.40	5.50	0.00
EWOC_NETS (wait, cohort size 3)	5.00	2.15	31.50	53.90	7.45	0.00
EWOC-NETS (no wait, cohort size 1)	1.72	2.60	43.05	47.22	5.42	0.00
EWOC-TITE	27.00	4.81	21.85	31.08	15.07	0.18
EWOC (wait, cohort size =1)	3.73	8.68	33.77	42.70	10.97	0.15
EWOC (wait, cohort size =3)	5.35	6.10	26.30	45.45	16.65	0.15
Over-Toxicity Scenario	Dose Level					
	1	**2**	3	4	5	6
Probability of DLT	0.08	**0.24**	0.33	0.44	0.56	0.76
ANETS	0.41	**0.49**	0.53	0.59	0.65	0.75
EWOC-NETS-TITE	5.27	40.92	42.02	9.70	0.1	0
EWOC-NETS (wait, cohort size 1)	9.53	46.35	37.73	6.37	1.67	0
EWOC_NETS (wait, cohort size 3)	12.40	41.60	37.85	8.15	0	0
EWOC-NETS (no wait, cohort size 1)	3.55	38.48	48.18	9.78	0	0
EWOC-TITE	31.70	34.00	21.27	12.18	0.85	0
EWOC (wait, cohort size =1)	22.83	42.97	24.50	9.22	0.48	0
EWOC (wait, cohort size =3)	24.20	42.20	23.95	9.10	0.55	0

Table 3.11 and Table 3.12 use other toxicity scenarios. In these two comparisons, we fixed the true MTD to dose level three with target toxicity level 33%. However, the TNETS

varies according to different toxicity profile. The toxicity profiles were described in previous section 3.4.2 in scenario 4 and scenario 5.

Table 3.11 MTD recommended for toxicity profile scenario 4 and scenario 5

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	0	10	65	25	0	0
EWOC-NETS (wait, cohort size 1)	0	5	67	28	0	0
EWOC_NETS (wait, cohort size 3)	0	4	75	21	0	0
EWOC-NETS (no wait, cohort size 1)	0	5	64	31	0	0
EWOC-TITE	22	13	42	23	0	0
EWOC (wait, cohort size =1)	1	14	65	20	0	0
EWOC (wait, cohort size =3)	1	18	58	23	0	0
Under-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.27	0.36	**0.41**	0.48	0.56	0.67
EWOC-NETS-TITE	0	9	69	22	0	0
EWOC-NETS (wait, cohort size 1)	0	5	73	22	0	0
EWOC_NETS (wait, cohort size 3)	0	8	61	31	0	0
EWOC-NETS (no wait, cohort size 1)	0	12	63	25	0	0
EWOC-TITE	22	13	42	23	0	0
EWOC (wait, cohort size =1)	1	14	65	20	0	0
EWOC (wait, cohort size =3)	1	18	58	23	0	0
Over-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.41	0.49	**0.53**	0.59	0.65	0.75
EWOC-NETS-TITE	0	8	67	24	1	0
EWOC-NETS (wait, cohort size 1)	0	3	75	22	0	0
EWOC_NETS (wait, cohort size 3)	0	12	63	25	0	0
EWOC-NETS (no wait, cohort size 1)	0	6	65	29	0	0
EWOC-TITE	22	13	42	23	0	0
EWOC (wait, cohort size =1)	1	14	65	20	0	0
EWOC (wait, cohort size =3)	1	18	58	23	0	0

Table 3.12 Patients' distribution in dose levels for scenario 4 and scenario 5

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	2.82	17.25	52.53	26.57	0.83	0.00
EWOC-NETS (wait, cohort size 1)	4.27	18.07	47.58	28.08	2.00	0.00
EWOC_NETS (wait, cohort size 3)	5.15	16.80	50.85	26.10	1.10	0.00
EWOC-NETS (no wait, cohort size 1)	3.27	10.92	60.72	24.97	0.13	0.00
EWOC-TITE	31.82	12.67	26.97	24.03	4.52	0.00
EWOC (wait, cohort size =1)	6.80	25.37	44.33	21.37	2.13	0.00
EWOC (wait, cohort size =3)	8.75	28.35	37.50	22.05	3.35	0.00
Under-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.27	0.36	**0.41**	0.48	0.56	0.67
EWOC-NETS-TITE	2.75	17.80	55.07	23.58	0.80	0
EWOC-NETS (wait, cohort size 1)	3.17	17.30	58.02	21.03	0.48	0
EWOC_NETS (wait, cohort size 3)	6.05	14.10	51.00	27.60	1.25	0
EWOC-NETS (no wait, cohort size 1)	1.85	15.62	58.13	24.37	3.33	0
EWOC-TITE	31.82	12.67	26.97	24.03	4.52	0
EWOC (wait, cohort size =1)	6.80	25.37	44.33	21.37	2.13	0
EWOC (wait, cohort size =3)	8.75	28.35	37.50	22.05	3.35	0
Over-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.41	0.49	**0.53**	0.59	0.65	0.75
EWOC-NETS-TITE	2.52	17.37	53.50	25.42	1.20	0
EWOC-NETS (wait, cohort size 1)	1.82	21.50	56.45	19.03	1.20	0
EWOC_NETS (wait, cohort size 3)	6.05	23.20	49.55	20.55	0.65	0
EWOC-NETS (no wait, cohort size 1)	1.83	22.18	54.75	21.08	0.15	0
EWOC-TITE	31.82	12.67	26.97	24.03	4.52	0
EWOC (wait, cohort size =1)	6.80	25.37	44.33	21.37	2.13	0
EWOC (wait, cohort size =3)	8.75	28.35	37.50	22.05	3.35	0

The following listed tables (Table 3.13 – Table 3.16) are the simulation results using sample size 30. Sample size 30 is usually the practical scenario for phase I clinical trials. We will also discuss the result in the subsequent section 3.5.

Table 3.13 Percent of MTD recommended according to ANETS = 47.6% sample of 30

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	0	3	49	47	1	0
EWOC-NETS (wait, cohort size 1)	0	6	63	31	0	0
EWOC_NETS (wait, cohort size 3)	0	5	57	38	1	0
EWOC-NETS (no wait, cohort size 1)	0	3	52	45	0	0
EWOC-TITE	31	10	31	25	3	0
EWOC (wait, cohort size =1)	0	16	55	27	1	0
EWOC (wait, cohort size =3)	1	18	51	27	2	0
Under-Toxicity Scenario	Dose Level					
	1	2	3	**4**	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.27	0.36	0.41	**0.48**	0.56	0.67
EWOC-NETS-TITE	0	0	16	74	9	0
EWOC-NETS (wait, cohort size 1)	0	1	28	67	5	0
EWOC_NETS (wait, cohort size 3)	0	1	26	67	6	0
EWOC-NETS (no wait, cohort size 1)	0	0	24	72	4	0
EWOC-TITE	32	2	13	38	15	0
EWOC (wait, cohort size =1)	0	2	25	62	11	0
EWOC (wait, cohort size =3)	0	2	30	56	12	0
Over-Toxicity Scenario	Dose Level					
	1	**2**	3	4	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.41	**0.49**	0.53	0.59	0.65	0.75
EWOC-NETS-TITE	0	42	48	10	0	0
EWOC-NETS (wait, cohort size 1)	0	47	46	7	0	0
EWOC_NETS (wait, cohort size 3)	0	42	48	10	0	0
EWOC-NETS (no wait, cohort size 1)	0	34	57	9	0	0
EWOC-TITE	30	26	34	9	2	0
EWOC (wait, cohort size =1)	1	50	46	3	2	0
EWOC (wait, cohort size =3)	7	41	45	7	0	0

Table 3.14 Percent of patients treated at each dose level sample of 30

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	3.37	5.97	36.71	48.63	3.37	5.97
EWOC-NETS (wait, cohort size 1)	4.63	15.51	51.27	27.96	4.63	15.51
EWOC_NETS (wait, cohort size 3)	10.42	12.44	49.10	27.46	10.42	12.44
EWOC-NETS (no wait, cohort size 1)	3.47	5.91	63.37	26.75	0.50	0.00
EWOC-TITE	33.75	8.59	24.04	26.65	6.95	2.00
EWOC (wait, cohort size =1)	10.05	23.51	38.36	24.11	3.96	0.00
EWOC (wait, cohort size =3)	14.58	17.22	32.84	29.20	6.16	0.00
Under-Toxicity Scenario	Dose Level					
	1	2	3	**4**	5	6
Probability of DLT	0.08	0.24	0.33	0.44	0.56	0.76
ANETS	0.27	0.36	0.41	**0.48**	0.56	0.67
EWOC-NETS-TITE	3.35	1.35	21.77	57.47	16.04	0.00
EWOC-NETS (wait, cohort size 1)	3.51	5.85	35.01	49.69	5.95	0.00
EWOC_NETS (wait, cohort size 3)	10.02	4.02	32.50	46.88	6.58	0.00
EWOC-NETS (no wait, cohort size 1)	3.42	2.61	43.39	46.46	4.12	0.00
EWOC-TITE	34.61	3.17	13.47	27.95	20.38	0.43
EWOC (wait, cohort size =1)	5.71	9.87	27.59	42.29	14.37	0.17
EWOC (wait, cohort size =3)	10.40	7.26	25.98	43.26	12.80	0.30
Over-Toxicity Scenario	Dose Level					
	1	**2**	3	4	5	6
Probability of DLT	0.08	**0.24**	0.33	0.44	0.56	0.76
ANETS	0.41	**0.49**	0.53	0.59	0.65	0.75
EWOC-NETS-TITE	4.39	14.21	46.86	33.25	1.30	0.00
EWOC-NETS (wait, cohort size 1)	8.83	28.94	48.90	13.08	0.25	0.00
EWOC_NETS (wait, cohort size 3)	12.48	23.24	50.18	13.96	0.14	0.00
EWOC-NETS (no wait, cohort size 1)	5.33	21.74	61.00	11.80	0.13	0.00
EWOC-TITE	35.93	14.95	24.77	22.36	1.98	0.00
EWOC (wait, cohort size =1)	22.54	33.49	29.81	13.25	0.91	0.00
EWOC (wait, cohort size =3)	22.50	27.66	32.88	15.76	1.20	0.00

Table 3.15 MTD recommended for toxicity profile scenario 4 and scenario 5 sample of 30

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	0	3	49	47	1	0
EWOC-NETS (wait, cohort size 1)	0	6	63	31	0	0
EWOC_NETS (wait, cohort size 3)	0	5	57	38	1	0
EWOC-NETS (no wait, cohort size 1)	0	3	52	45	0	0
EWOC-TITE	31	10	31	25	3	0
EWOC (wait, cohort size =1)	0	16	55	27	1	0
EWOC (wait, cohort size =3)	1	18	51	27	2	0
Under-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.27	0.36	**0.41**	0.48	0.56	0.67
EWOC-NETS-TITE	0	4	53	42	1	0
EWOC-NETS (wait, cohort size 1)	0	7	61	32	0	0
EWOC_NETS (wait, cohort size 3)	0	6	62	31	0	0
EWOC-NETS (no wait, cohort size 1)	0	5	55	39	1	0
EWOC-TITE	31	10	31	25	3	0
EWOC (wait, cohort size =1)	0	16	55	27	1	0
EWOC (wait, cohort size =3)	1	18	51	27	2	0
Over-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.41	0.49	**0.53**	0.59	0.65	0.75
EWOC-NETS-TITE	0	2	49	47	1	0
EWOC-NETS (wait, cohort size 1)	0	7	66	28	0	0
EWOC_NETS (wait, cohort size 3)	0	4	62	34	1	0
EWOC-NETS (no wait, cohort size 1)	0	7	56	36	0	0
EWOC-TITE	31	10	31	25	3	0
EWOC (wait, cohort size =1)	0	16	55	27	1	0
EWOC (wait, cohort size =3)	1	18	51	27	2	0

Table 3.16 Patients' distribution in dose levels for scenario 4 and scenario 5 sample of 30

Target Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.34	0.43	**0.48**	0.54	0.61	0.71
EWOC-NETS-TITE	3.37	5.97	36.71	48.63	3.37	5.97
EWOC-NETS (wait, cohort size 1)	4.63	15.51	51.27	27.96	4.63	15.51
EWOC_NETS (wait, cohort size 3)	10.42	12.44	49.10	27.46	10.42	12.44
EWOC-NETS (no wait, cohort size 1)	3.47	5.91	63.37	26.75	0.50	0.00
EWOC-TITE	33.75	8.59	24.04	26.65	6.95	2.00
EWOC (wait, cohort size =1)	10.05	23.51	38.36	24.11	3.96	0.00
EWOC (wait, cohort size =3)	14.58	17.22	32.84	29.20	6.16	0.00
Under-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.27	0.36	**0.41**	0.48	0.56	0.67
EWOC-NETS-TITE	3.37	5.79	37.40	49.09	4.35	0.00
EWOC-NETS (wait, cohort size 1)	4.99	16.37	49.97	27.71	0.96	0.00
EWOC_NETS (wait, cohort size 3)	10.82	12.62	49.40	26.68	0.48	0.00
EWOC-NETS (no wait, cohort size 1)	3.71	7.65	62.93	25.23	0.47	0.00
EWOC-TITE	33.75	8.59	24.04	26.65	6.95	2.00
EWOC (wait, cohort size =1)	10.05	23.51	38.36	24.11	3.96	0.00
EWOC (wait, cohort size =3)	14.58	17.22	32.84	29.20	6.16	0.00
Over-Toxicity Scenario	Dose Level					
	1	2	**3**	4	5	6
Probability of DLT	0.08	0.24	**0.33**	0.44	0.56	0.76
ANETS	0.41	0.49	**0.53**	0.59	0.65	0.75
EWOC-NETS-TITE	3.41	4.23	39.97	47.50	4.89	0.00
EWOC-NETS (wait, cohort size 1)	4.53	16.79	51.51	26.59	0.57	0.00
EWOC_NETS (wait, cohort size 3)	10.92	11.56	49.82	26.82	0.88	0.00
EWOC-NETS (no wait, cohort size 1)	3.89	9.00	60.35	26.35	0.40	0.00
EWOC-TITE	33.75	8.59	24.04	26.65	6.95	2.00
EWOC (wait, cohort size =1)	10.05	23.51	38.36	24.11	3.96	0.00
EWOC (wait, cohort size =3)	14.58	17.22	32.84	29.20	6.16	0.00

3.4.2 Comparison of Trial Duration

Trial duration is the other area of interest and also the motivation for this part of research study. We construct two comparison scenarios: (1) We use different inter-patient mean arrival time, which represents the average time to recruit a new patient. An exponential distribution was used to generate the arrival time. The reason we chose to use exponential distribution rather than

Poisson is: arrival time generated by exponential distribution is more spread than the one generated by Poisson, which is more appropriate and closer to the scenarios of practical clinical trial experiment. (2) We use different sample sizes for simulating the trial durations, because also interested in how the sample size affects the trial duration. Table 3.17 and Table 3.18 displays the simulated phase I trial durations.

Table 3.17 Simulated trial durations by different sample size

	Sample Size		
Design of Phase I Trials	15	30	60
EWOC-NETS-TITE	139	254	477
EWOC-TITE	138	261	469
EWOC-NETS (no wait)	141	249	476
EWOC-NETS (wait, cohort size 1)	428	848	1,688
EWOC (wait, cohort size 1)	416	854	1,653
EWOC-NETS (wait, cohort size 3)	177	322	603
EWOC (wait, cohort size 3)	184	316	612

Table 3.18 Simulated trial duration by mean arrival time

	Mean Inter-Patient Arrival Time		
Design of Phase I Trials	7	28	100
EWOC-NETS-TITE	254	882	2,999
EWOC-TITE	260	877	3,000
EWOC-NETS (no wait)	254	875	3,000
EWOC-NETS (wait, cohort size 1)	848	979	3,005
EWOC (wait, cohort size 1)	852	977	3,001
EWOC-NETS (wait, cohort size 3)	322	883	3,000
EWOC (wait, cohort size 3)	329	884	2,999

3.5 Conclusion and Discussion

We will discuss the above listed results regarding (1) accuracy of MTD estimation, (2) simulated total days to complete the trial, (3) model performance with different toxicity information. When we first initialize the idea to design a new phase trial, it is actually the second topic we discussed motivated us to seek for a solution. We then looked for a solution for the practice that cancer patients cannot bear to stay in an experiment without any treatment intervention. As long as we started the research study, more issues have been raise out, such as

how to fully utilize all information, since every piece of information in a cancer clinical trial is quite precious. Many concerns were proposed first, and then is the process of seeking solutions. After comparing many popularly used designs for cancer phase I trials, we decided to choose EWOC as our primary model. The essential thinking is: EWOC has a mechanism to control the probability of overdosing a patient. Our way of thinking is somehow down to the earth; the safety issue is crucial and needs to be put into the highest priority. Therefore, the hybrid design comes out with all these concerns.

Now, let's examine the performance of the hybrid design, and explore whether this design is appropriate and what is the further research direction starts from here.

3.5.1 Comparison of MTD Accuracy

The design performance is quite solid for MTD estimation when we examine the results from Table 3.9 – Table 3.12. In most scenarios, the proposed hybrid design can accurately detect the true MTD, especially when the dose is predefined to a moderate level with an equal ratio in both non-DLT and DLT categories. Use Table 3.9 as an example, and examine the designs in a pairwise manner.

(1) We review the results of the basic EWOC design. Both of them provide accurate MTD estimations while a better estimation is given if the cohort size is 1. From the design scheme, we know that the new MTD estimation is updated after complete observation (in an assessment window) of each cohort. When the sample size is fixed, a smaller cohort size means more frequent update. This would be a possibility that explains the percentage in target scenario. But in under-toxicity and over toxicity scenario, the result is opposite. So there may be another underlying reason. A possible explanation could be: when using larger cohort, the dose-toxicity relationship tends to more stable, because more patients were observed, and the random variation

decreased. Therefore, when cohort size is 1, the estimate result at some extend depends on how "normal" the patient's reaction is. Furthermore, we check all EWOC results in the following tables. It is really random that which estimate is better—cohort size 1 or 3. Hence, the difference in estimation accuracy according to cohort size can be offset. There is no significant distinction between it.

(2) We go a step further; compare the result of EWOC and EWOC-TITE. Because the characteristic of time-to-event approach is to enroll a patient in a staggered fashion, we will compare it with EWOC with cohort size 1. Although EWOC-TITE could identify the true MTD, simulation results show that there is more than half chance that the true MTD would not be identified. The advantage of "completeness" is of great impression in this result. But there is another possibility that the lower percent dues to the selection of mean arrival time. In these tables, the mean arrival time we used is 7 days, which means the assessment interval between patients is very small that conducts to a very small weight induced in the model. When weight and information obtained sometimes is too small, it may confound the results by leading the results to another direction. So in further research, we would like to explore more scenarios in inter-patient arrival time.

(3) We compare EWOC and EWOC-NETS. It can be easily concluded that the EWOC-NETS is superior to EWOC. With the same waiting fashion, EWOC-NETS has a higher probability to detect the true MTD. This result is under our expectation, because EWOC-NETS takes advantage of both "completeness" and "more informative". Suppose the difference in percentage dues to the adoption of NETS, it accounts for as many as 10% differences in MTD estimation. It made us reflect the goal of the phase I clinical trial, which is to establish the dose-toxicity relationship as well as seek for an optimal dose. It at some extends shows how important

and how precious toxicity information is in phase I trials. So it is definitely crucial to use as much toxicity information as possible.

(4) In addition, we compare EWOC-NETS between waiting and no-waiting fashion. The result shows that omitting incomplete observation does not affect the estimation result as much as we expected. From the other point of view, it confirms the conclusion of (3) that, toxicity information could a key for further improvement in effectiveness in phase I trials.

(5) Finally, we compare accuracy of MTD estimation of our proposed hybrid design with EWOC, EWOC-NETS, and EWOC-TITE. Without any doubt, EWOC-NETS has the highest accuracy, followed closely by our proposed hybrid design — EWOC-NETS-TITE. If expressed in an inequality, it can be considered that the accuracy in MTD estimation is: EWOC-NETS > EWOC-NETS-TITE > EWOC. The "incompleteness" in observation is partially made up by the time-to-event approach and partially replenished by using more toxicity information. However, in the overall evaluation, EWOC-NETS-TITE has a unique advantage that the trial duration is significantly reduced.

(6) However, when encounter the over-toxicity scenario all EWOC-NETS derived designs tend to overestimate the MTD, which could be considered a little aggressive in contrary to the basic conservative design EWOC. One of the possibilities is that the overestimating may due to the choice of toxicity profile. In phase I clinical trials, expert inputs from clinicians with abundant experience and well understanding of the testing drug is an essential component. And this issue would need further investigation in the later research.

Another factor to evaluate the model performance is the percent of patients treated at each dose level. Since the guiding principle in phase I clinical trials is to treat as many patients as possible to therapeutic dose levels and avoiding exposures to lower dose levels. The results

shown were pretty promising that most patients were treated at the target dose levels. The only outlier is EWOC-TITE methods. The reason leads to the result is: when the accrual rate is relative high, many patients were enrolled in the trial during the first couple of days. Then all observations are incomplete, although a time-to-event approach was used. But too little information can be used to update the dose level.

3.5.2 Comparison of Trial Length

Table 3.17 and Table 3.18 illustrate the trial durations using different sample sizes and by different mean arrival time. In these two tables, we can consider the nature of EWOC-NETS-TITE, EWOC-TITE, and EWOC-NETS no-waiting fashion is same with respect to trial duration. Since all three designs allow newly recruited patients to enter the trial immediately. The EWOC-NETS-TITE can effectively reduce the trial duration comparing to cohort size 1 EWOC derived designs, and even cohort size 3 EWOC derived designs. As is the common scenario that patients are treated in a small group, like 3 patients per group. The EWOC-NETS-TITE could dramatically reduce the length by more than 20% without sacrificing any accuracy in MTD estimation. Therefore, EWOC-NETS-TITE could be considered as an effective design for phase I cancer trials.

4 PERSONALIZED MTD FOR PHASE I CLINICAL TRIALS

The recent development in technology improves our understanding in cancer biology and drug metabolism that prompts the emergence of personalized medicine. Personalized medicine accommodates individual patients' needs and differences in drug tolerance. This method could estimate the dose level given to patients with respect to patients' characteristics. In this chapter, we will extend our proposed phase I design — EWOC-NETS-TITE to take into account patients' baseline covariates in order to recommend a dose for subsequent phase II trial. We will provide a description of the approach that utilizes patient pretreatment characteristics in terms of covariate factor to improve the efficiency with the estimation of MTD. The design of the trial permitted a continual adjustment of the model used to tailor the dose to each patient's individual need. The performance of the design will be evaluated with extensive simulations by comparing models with respect to the accuracy and efficiency of the estimate of the conditional MTD. We will also discuss the potential applications of the design to incorporate patient genomic information and associations with toxicity profiles.

4.1 Introduction

One of the primary goals of early-phase cancer clinical trials is to determine the optimal dose of a new drug to take forward into subsequent, outcome-oriented clinical trials. Usually, once a dose is scheduled for follow up phase II and phase III trials, we rarely go back to explore alternatives. Throughout history, the practice of medicine has largely been reactive. Even today, we have to wait until the onset of diseases and then try to treat or cure them due to only little was known to the genetic and environmental factors. However, these factors have caused major diseases such as cancer, Alzheimer's and diabetes, and so on; but our efforts to treat them are often imprecise, unpredictable and ineffective (Offit, 2011). Without taking into account of these

factors, the same dose is then allocated all patients, adjustments only according to the traditional mg/kg or similar dosage approaches. In other words, the new drugs or treatments we devise are tested on broad populations and are prescribed using statistical averages, for example, a group of patients with same the disease usually receive the same average dosage (Zhao, et al., 2013). However, the drugs on the market only work for half of those who take it (Figure 4-1 demonstrate the heterogeneity of drug effects among patient group). It is extremely low for anti-cancer drugs, about 25% (Arienti, et al., 2011). It is the time to refine our understanding of how to choose the right dose and schedule for a given patient with specific covariate. The motivation behind is: because in today's world, finding a small benefit among a large patient population is often a big step forward as drug development is such a complicated process. And the small benefit is often enough to obtain regulatory approval and to influence equally important practice guidelines. Usually, MTD estimation is based on the assumption that more is better. Higher dose intensity means more chances to be able to cure the patient. Most cancer treatments are dosed according to this principle, however more and more variations in treatment effects were found among patients with the same dose levels.

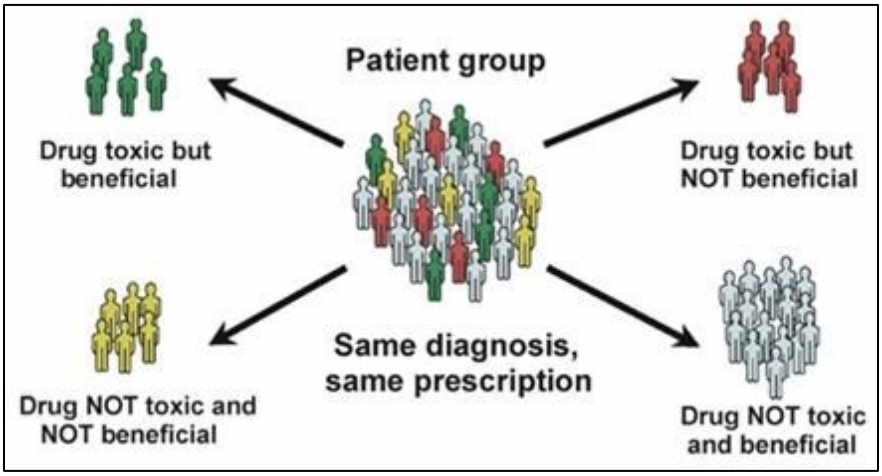

Figure 4-1Heterogeneity among patients

The reason for this heterogeneity is that these patients have different genetics and environmental profiles that include demographic characteristics, concomitant diseases, and genetics, and so on. Both genetic factors and environmental factors interplay to affect therapeutic effect of treatment or drug (Chan, et al., 2011). With recent experience and understanding of cancer biology, we have learned that cancer is a much more complex disease than we previously thought. Each patient is distinct from other patients with respect to clinical activity, prognosis, tumor response and tolerance to treatment in addition to differences in risk of recurrence, second malignancy and long-term complications of treatment (Fleck, et al., 2012). Therefore, development of personalized medicine is more important for cancer patients than for those other disease due to its heterogeneity among patients. Personalized medicine has emerged as an advance approach to achieve optimal medical outcomes in context of patient's genetic and environmental profile. At present, personalized medicine has become a trend and a direction for future endeavor in the field of medicine.

The primary goal of a phase I trial is to estimate the MTD under safe administration of the new drug or treatment under development. Normally, a cohort of patients is treated at the same dose of a drug, some may experience toxicities, or even experience of DLT, but the others may experience none. The question is often asked: why the responses differ so much? The main reason is the heterogeneity of patients. Patients' genetics and environmental profiles have been found to affect their acceptability to drug in term of dosage. Known reasons include patient's vulnerability to an exaggerated pharmacodynamics, differences in genetic susceptibility, and drug-to-drug interactions (Babb, et al., 2001). In order to achieve an optimal therapeutic effect of a drug for every patient, a personalized MTD of the drug for individual patient may be the answer.

Since rule based phase I designs or nonparametric phase I designs cannot estimate the MTD with in terms of covariates due to their algorithms. Hence, we choose to use parametric model to incorporate patients' characteristics. In order to use as much information as possible, we will use our proposed hybrid design to accommodate the covariate information. A binary covariate approach has been proposed for EWOC by Tighiouart *et al.* shows that improvement in probability of overdosing a patient when there is a significant baseline binary covariate (Tighiouart, et al., 2012). In this dissertation, we will use our proposed hybrid design EWOC-NETS-TITE as a base model, and incorporate a binary covariate into the model. In addition, we will compare for more general scenario by using a continuous covariate for patient's characteristics.

4.2 Personalized MTD for Phase I Clinical Trials

In chapter 3, we have provided a description of our hybrid design — EWOC-NETS-TITE for estimation of MTD which allows us to fully utilize all toxicity information and take into account of time factor for dose finding with overdose control. In this chapter, we will accommodate this design to the patient's covariate information through dose-toxicity relationship. Follows the assumption of previous chapter, a logistic model is adopted for the basic dose-toxicity relationship. But within the new dose-toxicity relationship, we assume the probability of DLT is a function of both factors: dose level allocated to patient and patient's self-characteristics. z is used to represent the patient's own characteristics which measures the physical differences between patients by integrate information from specific aspects. Other notations have the same meaning as used in previous chapter.

4.2.1 Binary Covariate Effect

Binary variables are often used to describe patients' characteristics and classify patients into subgroups. These variables cannot be replaced with continuous values as their practical meanings will be missing. Such patients' characteristics include gender, ethnicity, diabetes (Yes or No), and source of primary care — hospital versus home, and so on. Hence, in this section, we will incorporate patients' binary characteristics as a source of effect for the toxicity response. EWOC-NETS-TITE will be our base model. It is obvious that using the binary characteristic patients are divided into two groups. Here we will provide a description on how to derive the MTD for each group of patients.

Assume γ_0, γ_1 are the MTD for the patients in different groups divided by the binary characteristic. We use z denote the covariate, then $z = 0, 1$ and

$$\gamma_0 = \text{MTD} \mid z = 0$$

$$\gamma_1 = \text{MTD} \mid z = 1$$

The NETS for both groups of patients are defined as:

$$s_0(x) = \text{NETS} \mid \text{Dose} = x, z = 0$$

$$s_1(x) = \text{NETS} \mid \text{Dose} = x, z = 1$$

where standardized dose level $x \in [x_{\min}, x_{\max}] = [0,1]$. A logistic model is also used to establish the dose-toxicity relationship:

$$s_0(x) = F(\alpha + \beta x + \delta \cdot 0) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

$$s_1(x) = F(\alpha + \beta x + \delta \cdot 1) = \frac{\exp(\alpha + \beta x + \delta)}{1 + \exp(\alpha + \beta x + \delta)}$$

Thus the initial probabilities of getting target NETS can also be specified with respect to the covariate effect as:

$$\rho_0 = s_0(x_{\min}) = \frac{\exp(\alpha + \beta x_{\min})}{1 + \exp(\alpha + \beta x_{\min})}$$

$$\rho_1 = s_1(x_{\min}) = \frac{\exp(\alpha + \beta x_{\min} + \delta)}{1 + \exp(\alpha + \beta x_{\min} + \delta)}$$

Then the target NETS is associated with γ_0 and γ_1 and can be expanded as:

$$\tilde{\theta} = \frac{\exp(\alpha + \beta\gamma_0)}{1 + \exp(\alpha + \beta\gamma_0)} = \frac{\exp(\alpha + \beta\gamma_1 + \delta)}{1 + \exp(\alpha + \beta\gamma_1 + \delta)}$$

Here $\beta \geq 0$ that implies the non-decreasing dose-toxicity relationship. This model also implies a constant odds ratio of toxicity between the two groups of patients in the sense that this odds ratio does not depend on the dose level.

$$\frac{\rho_0/(1-\rho_0)}{\rho_1/(1-\rho_1)} = \frac{\exp(\alpha + \beta x_{\min})}{\exp(\alpha + \beta x_{\min} + \delta)} = \exp(-\delta)$$

We can derive the parameters in the model in terms of γ_0, $\tilde{\theta}$, ρ_0 and ρ_1 as:

$$\alpha = \frac{\gamma_0 \text{logit}(\rho_0) - x_{\min}\text{logit}(\tilde{\theta})}{\gamma_0 - x_{\min}}$$

$$\beta = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_0)}{\gamma_0 - x_{\min}}$$

$$\delta = \text{logit}(\rho_1) - \text{logit}(\rho_0)$$

After specifying the prior distribution that $(\rho_1, \rho_2) \sim U[0, \tilde{\theta}] \times U[0, \tilde{\theta}]$ and $\gamma_0 \sim U[x_{\min}, x_{\max}]$ for parameters ρ_1, ρ_2 and γ_0, we denote $\pi_{z,n}(\gamma)$ as the marginal posterior cumulative density function of γ given D_n with different covariate effects $z = 0$ or $z = 1$, then the $(n+1)^{th}$ patient receives the dose

$$x_{c,n+1} = \pi_{z,n}^{-1}(\alpha')$$

so that the posterior probability of exceeding the MTD is equal to the feasibility bound α'. Let

$$\lambda_i(x_i, z_i) = \alpha + \beta x_i + \delta z_i$$

$$= \frac{\gamma_0 \text{logit}(\rho_2) - x_{\min}\text{logit}(\tilde{\theta})}{\gamma_0 - x_{\min}} + \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_0)}{\gamma_0 - x_{\min}} x_i + [\text{logit}(\rho_1) - \text{logit}(\rho_0)]c_i$$

Thus, the likelihood taking into account time factor w_i of the data $D_j = \{(x_i, s_i, z_i); i = 1, \dots, k\}$ in terms of parameters γ_0, ρ_0, ρ_1 is

$$L(\rho_0, \rho_1, \gamma_0 | D_j)$$

$$= \prod_{i=1}^{j} w_i^{s_i} \cdot \frac{\ln\left[\{1 - w_i \exp(\lambda_i(x_i, z_i))\}^{-1} - 1\right]}{(2w_i - 1) \cdot \exp(\lambda_i(x_i, z_i)) - 1}$$

$$\times \left[\exp(\lambda_i(x_i, z_i))\right]^{s_i} \times \left[1 + (1 - w_i) \cdot \exp(\lambda_i(x_i, z_i))\right]^{1 - s_i}$$

Using $h(\rho_0, \rho_1, \gamma_0)$ denote the prior distribution on $[0, \tilde{\theta}] \times [0, \tilde{\theta}] \times [x_{\min}, x_{\max}]$, then the posterior distribution of $(\rho_0, \rho_1, \gamma_0)$ is

$$\pi(\rho_1, \rho_2, \gamma_{\max} | D_j) = \frac{L(\rho_0, \rho_1, \gamma_0 | D_j) \cdot h(\rho_0, \rho_1, \gamma_0)}{\iiint_{[0,\tilde{\theta}] \times [0,\tilde{\theta}] \times [x_{\min}, x_{\max}]} L(\rho_0, \rho_1, \gamma_0 | D_j) h(\rho_0, \rho_1, \gamma_0) d\rho_0 d\rho_1 d\gamma_0}$$

Once the posterior $\pi(\rho_0, \rho_1, \gamma_0 | D_j)$ has been obtained, the relationship

$$\gamma_z = \frac{\text{logit}(\tilde{\theta}) - \alpha}{\beta} - \frac{\delta}{\beta} \cdot z$$

where

$$\frac{\delta}{\beta} = \ln\left[\frac{\rho_1(1 - \rho_0)}{\rho_0(1 - \rho_1)}\right] \ln\left[\frac{\tilde{\theta}(1 - \rho_0)}{\rho_0(1 - \tilde{\theta})}\right]^{-1} \gamma_0$$

can be used to derive $\pi_z(\cdot | D_j)$, the marginal posterior distribution of the MTD γ_0, γ_1 $(z = 0,1)$ given D_j.

4.2.2 Continuous Covariate Effect

In other occasions, it is more convenient to describe patients' characteristics using continuous variables which allow us to use as much information as possible. The advantage of

using binary variables is its simplicity, however, the cutoff criteria to define the binary variable is subjective at some extent; for example, a patient is considered to have high blood pressure. Instead, it is more informative to use his/her blood pressure directly other than the high/low result. And the MTD estimated based on the continuous covariate could be more accurate and targeted for individual patient.

Assume γ_z is the MTD for the patient with a continuous covariate z.

$$\gamma_z = \text{MTD} \mid Z = z$$

The NETS for patient is defined as:

$$s_z(x) = \text{NETS} \mid \text{Dose} = x, Z = z$$

where z and dose level x are both standardized; z is bounded with $[z_1, z_2] = [0,1]$, and dose level $x \in [x_{\min}, x_{\max}] = [0,1]$. A logistic model is also used to establish the dose-toxicity relationship:

$$s_z(x) = F(\alpha + \beta x + \delta z)$$

$$= \frac{\exp(\alpha + \beta x + \delta z)}{1 + \exp(\alpha + \beta x + \delta z)}$$

We use $\gamma_{\max}$ to denote the MTD when patient's maximum covariate effect is identified, that is:

$$\gamma_{\max} = \gamma_{z_2} = \text{MTD} \mid z = 1$$

$$\gamma_{\min} = \gamma_{z_1} = \text{MTD} \mid z = 1$$

And initial probabilities of getting target NETS can also be specified with respect to the covariate effect as:

$$\rho_1 = s_{z_1}(x_{\min})$$

$$\rho_2 = s_{z_2}(x_{\min})$$

Then the target NETS is associated with $\gamma_{\max}$ and expanded as:

$$\tilde{\theta} = \frac{\exp(\alpha + \beta\gamma_{\max} + \delta z_2)}{1 + \exp(\alpha + \beta\gamma_{\max} + \delta z_2)}$$

In addition, ρ_1 and ρ_2 are:

$$\rho_1 = \frac{\exp(\alpha + \beta x_{\min} + \delta z_1)}{1 + \exp(\alpha + \beta x_{\min} + \delta z_1)}$$

$$\rho_2 = \frac{\exp(\alpha + \beta x_{\min} + \delta z_2)}{1 + \exp(\alpha + \beta x_{\min} + \delta z_2)}$$

Here $\beta \geq 0$ that implies the non-decreasing dose-toxicity relationship. This model also implies a constant odds ratio of toxicity between the two patients in the sense that this odds ratio does not depend on the dose level.

$$\frac{\rho_1/(1-\rho_1)}{\rho_2/(1-\rho_2)} = \frac{\exp(\alpha + \beta x_{\min} + \delta z_1)}{\exp(\alpha + \beta x_{\min} + \delta z_2)} = \exp(\delta(z_1 - z_2))$$

We can derive the parameters in the model in terms of $\gamma_{\max}$, $\tilde{\theta}$, ρ_1 and ρ_2 as:

$$\alpha = \text{logit}(\rho_2) - x_{\min}\frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_2)}{\gamma_{\max} - x_{\min}} - z_2\frac{\text{logit}(\rho_1) - \text{logit}(\rho_2)}{z_1 - z_2}$$

$$\beta = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_2)}{\gamma_{\max} - x_{\min}}$$

$$\delta = \frac{\text{logit}(\rho_1) - \text{logit}(\rho_2)}{z_1 - z_2}$$

After specifying the prior distribution that $(\rho_1, \rho_2) \sim U[0, \tilde{\theta}] \times U[0, \tilde{\theta}]$ and $\gamma_{\max} \sim U[x_{\min}, x_{\max}]$ for parameters ρ_1, ρ_2 and $\gamma_{\max}$, we denote $\pi_{z,n}(\gamma)$ as the marginal posterior cumulative density function of γ given D_n, then the $(n+1)^{th}$ patient receives the dose

$$x_{z,n+1} = \pi_{z,n}^{-1}(\alpha')$$

so that the posterior probability of exceeding the MTD is equal to the feasibility bound α'. Let

$$\lambda_i(x_i, z_i) = \alpha + \beta x_i + \delta z_i$$

$$= \text{logit}(\rho_2) + (x_i - x_{\min})\frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_2)}{\gamma_{\max} - x_{\min}} - (z_2 - z_i)\frac{\text{logit}(\rho_1) - \text{logit}(\rho_2)}{z_1 - z_2}$$

Thus, the likelihood taking into account of time factor w_i of the data $D_j = \{(x_i, s_i, z_i); i = 1, \dots, k\}$ in terms of parameters $\gamma_{\max}, \rho_1, \rho_2$ is

$$L(\rho_1, \rho_2, \gamma_{\max}|D_j)$$

$$= \prod_{i=1}^{j} w_i^{s_i} \cdot \frac{\ln\left[\{1 - w_i \exp(\lambda_i(x_i, z_i))\}^{-1} - 1\right]}{(2w_i - 1) \cdot \exp(\lambda_i(x_i, z_i)) - 1}$$

$$\times \left[\exp(\lambda_i(x_i, z_i))\right]^{s_i} \times \left[1 + (1 - w_i) \cdot \exp(\lambda_i(x_i, z_i))\right]^{1 - s_i}$$

Using $h(\rho_1, \rho_2, \gamma_{\max})$ denote the prior distribution on $[0, \tilde{\theta}] \times [0, \tilde{\theta}] \times [x_{\min}, x_{\max}]$, then the posterior distribution of $(\rho_1, \rho_2, \gamma_{\max})$ is

$$\pi(\rho_1, \rho_2, \gamma_{\max}|D_j) = \frac{L(\rho_1, \rho_2, \gamma_{\max}|D_j) \cdot h(\rho_1, \rho_2, \gamma_{\max})}{\iiint_{[0,\tilde{\theta}]\times[0,\tilde{\theta}]\times[x_{\min}, x_{\max}]} L(\rho_1, \rho_2, \gamma_{\max}|D_j) h(\rho_1, \rho_2, \gamma_{\max}) d\rho_1 d\rho_2 d\gamma_{\max}}$$

Once the posterior $\pi(\rho_1, \rho_2, \gamma_{\max}|D_j)$ has been obtained, the relationship

$$\gamma_z = \frac{\text{logit}(\tilde{\theta}) - \alpha}{\beta} - \frac{\delta}{\beta} \cdot z$$

where

$$\frac{\delta}{\beta} = \ln\left[\frac{\rho_1(1 - \rho_2)}{\rho_2(1 - \rho_1)}\right] \ln\left[\frac{\tilde{\theta}(1 - \rho_2)}{\rho_2(1 - \tilde{\theta})}\right]^{-1} \gamma_{\max}$$

$$\frac{\text{logit}(\tilde{\theta}) - \alpha}{\beta} = \ln\left[\frac{\tilde{\theta}\rho_1(1 - \rho_2)^2}{(1 - \tilde{\theta})(1 - \rho_1)\rho_2^2}\right] \ln\left[\frac{\tilde{\theta}(1 - \rho_2)}{(1 - \tilde{\theta})\rho_2}\right]^{-1} \gamma_{\max}$$

can be used to derive $\pi_z(\cdot\,|D_j)$, the marginal posterior distribution of the MTD γ_z given D_j.

4.3 Simulation Study

A simulation study is conducted to evaluate the performance of our proposed phase I design for personalized MTD. The performance of the model can be evaluated from two aspects: the accuracy of MTD estimation, and the probability of overdosing a patient. The comparison will be made between model with consideration of patients' covariates and model without consideration of patients' covariates. A replicate of 1,000 trials was done for both part of the simulation study.

4.3.1 Simulation Plan

As discussed above, we proposed a very simple linear model to formulate the patient's covariate effect either binary or continuous. The estimations of the personalized MTD can be derived from the dose-toxicity relationship which is a function of the patient's covariate. That is, a larger absolute value of covariate coefficient $k = \frac{\delta}{\beta}$ relates to stronger effect, while values close or equal to zero implies patient's characteristic has little or even no impact on the toxicity response.

Simulation study will help us find evidence that include covariate effect will improve accuracy of MTD estimation. Hence, with the above assumption, we will plan the simulation study with two scenarios: (1) suppose patient's characteristics have impact on the toxicity response, which implies a nonzero value of $k = \frac{\delta}{\beta}$; (2) assume no covariate effect exists, which means $k = \frac{\delta}{\beta}$ is zero.

As shown above, the personalized MTD can be derived from the relationship

$$\gamma_z = \frac{\text{logit}(\tilde{\theta}) - \alpha}{\beta} - \frac{\delta}{\beta} \cdot z$$

once the posterior $\pi_z(\cdot|D_j)$ is obtained and all parameters have been updated. Therefore, simulation study of the first scenario is used to answer the questions: whether our model can correctly identify the covariate effect, as well as how accurate the estimations of both MTD and covariate effects are. When the covariate effect exists, it is easy to derive a range for the MTD according to different covariate effects. In contrary, a point estimate of MTD is the result for no effect assumption. Simulation study will be conducted to model both proposed covariate — binary and continuous variable.

(1) Using binary covariate assumption: for the first scenario, $k = \frac{\delta}{\beta}$ is predefined with a value in [0,1] represents the linear effect that standardized patient's covariate has on MTD before we start the simulation study. Meanwhile, we will provide a pre-specified value of ρ_0, which is the probability of TNETS for one group of patients with covariate $z = 0$. MTD γ_0 that associates with zero covariate group patients is also given. With all three parameters, we can derive an initial value for ρ_1. Using the initial guesses of $(\rho_0, \rho_1, \gamma_0)$, the equation of γ_z in terms of z $(z = 0, 1)$can be expressed with estimates of $\alpha, \beta,$ and δ after all patients finished their follow ups in the trial. For the second scenario, $k = \frac{\delta}{\beta}$ is predefined as zero meaning no covariate effect in reality. Similar to previous scenario, pre-specify γ_0, ρ_0 before starting the trial. Since no covariate effect is assumes, then ρ_0 and ρ_1 actually have the same clinical meaning and value. The personalized MTD could be established with parameters updated after completion of the designed simulation.

(2) Using continuous covariate assumption: for the first scenario, we will provide a predefined value for $k = \frac{\delta}{\beta}$ that ranges on [0,1] implies the linear effect that standardized patient's covariate has on MTD. Expert input of the probability of getting a target NETS for

patient with maximum covariate z_2 who is given the minimum dose is also determined. This ρ_2 is set up as the initial value in the Bayesian process. Whereas the MTD associates with maximum covariate $\gamma_{\max}$ is also pre-specified. With all three parameters, we can derive an initial guess for ρ_1. Using the initial guesses of $(\rho_1, \rho_2, \gamma_{\max})$, the equation of γ_z in terms of z can be expressed with estimates of α, β, and δ after all patients finished their follow ups in the trial. For the second scenario, $k = \frac{\delta}{\beta}$ is predefined as zero represents no linear effect that standardized patient's covariate has on MTD before we start the simulation study. Similar to previous scenario, pre-specify ρ_2 before starting the trial. Since no covariate effect is assumes, then ρ_1 and ρ_2 actually have the same clinical meaning. The personalized MTD could be established with parameters updated after completion of the designed simulation.

In order to detect the accuracy of our MTD estimation, two separate trial designs will be utilized. Other than the phase I design considering covariate effect, we will use EWOC-NETS-TITE as a "control design" ignoring the covariate effect. In addition to solely estimate the MTD, we will provide a comparison of the probability of overdosing a patient for each design, which is a very important criterion for evaluating the design. We will provide a detailed description of the simulation setup and process in section 4.3.2.

4.3.2 Simulation Setup

For simplicity and without loss of generality, z_i — patient's covariate, γ_z — personalized MTD, and x_i — dose level are standardized and bounded on [0,1]. Then we use the equation $\gamma_z = \frac{\text{logit}(\tilde{\theta}) - \alpha}{\beta} - \frac{\delta}{\beta} \cdot z$ as a general basis for the simulation study for both binary and continuous covariate. To keep consistency, the target NETS $\tilde{\theta}$ is still set up as 0.476 in all simulated trials.

(1) For the binary covariate assumption: with predefined γ_0 ρ_0 and covariate coefficient $k = \frac{\delta}{\beta}$, we can easily derive ρ_1 with as follows: given $x_{\min} = 0, x_{\max} = 1$, then

$$\alpha = \text{logit}(\rho_0)$$

$$\beta = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_0)}{\gamma_0}$$

$$\delta = \text{logit}(\rho_1) - \text{logit}(\rho_0)$$

Hence, γ_0 which represents the MTD of patient group with covariate $z = 0$ is:

$$\gamma_0 = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_0)}{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_1)}$$

Whereas the MTD of patient group with covariate $(z = 1)$ is similarly derived as:

$$\gamma_1 = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_0)}{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_1)} - k$$

The initial guesses for parameters $(\rho_0, \rho_1, \gamma_0)$ and other pre-specified model properties for different scenarios are listed below.

The first scenario assumes that covariate has impact on toxicity response:

Table 4.1 Model setup for MTD estimation with covariate effect

Intercept	Slope	Target NETS (%)	MTD at max. Cov.	Proportion of TNETS for min. Cov.	Proportion of TNETS for max. Cov.
$\frac{\text{logit}(\tilde{\theta}) - \alpha}{\beta}$	$k = \frac{\delta}{\beta}$	$\tilde{\theta}$	γ_0	ρ_0	ρ_1
0.5879	0.1879	47.60	0.5879	0.12	0.20
0.6875	0.2875	47.60	0.6875	0.09	0.20
0.7995	0.1995	47.60	0.7995	0.14	0.20
0.8007	0.2007	47.60	0.8007	0.25	0.30
0.8883	0.2883	47.60	0.8883	0.23	0.30
0.8819	0.2818	47.60	0.8819	0.12	0.20

While for the second scenario we assume there is no covariate effect in the model of dose-toxicity relationship:

Table 4.2 Model setup for MTD estimation without covariate effect

TNETS (%)	MTD for patient group with $z = 1$	MTD for patient group with $z = 0$	Proportion of TNETS for $z = 0$	Proportion of TNETS for $z = 1$
$\tilde{\theta}$	γ_1	γ_0	ρ_0	ρ_1
47.60	0.40	0.40	0.20	0.20
47.60	0.40	0.40	0.30	0.30
47.60	0.60	0.60	0.20	0.20
47.60	0.60	0.60	0.30	0.30

We use the MCMC sampling method to simulate the Bayesian framework of the three parameters $(\rho_0, \rho_1, \gamma_0)$ of interest. The assessment time window was set up to 28 days. Time and events were simulated using an exponential distribution with mean patient's arrival time of 7 days.

(2) For the continuous covariate assumption: with predefined $\gamma_{\max}$, ρ_2 and covariate coefficient $k = \frac{\delta}{\beta}$, we can easily derive ρ_1 with as follows: given $z_1 = 0,\ z_2 = 1, x_{\min} = 0, x_{\max} = 1$, then

$$\alpha = \text{logit}(\rho_1)$$

$$\beta = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_2)}{\gamma_{\max}}$$

$$\delta = \text{logit}(\rho_2) - \text{logit}(\rho_1)$$

Hence, $\gamma_{\max}$ which represents the MTD of maximum patient's covariate is:

$$\gamma_{\max} = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_1)}{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_2)} - k$$

Whereas the MTD of minimum patient's covariate is similarly derived as:

$$\gamma_{\min} = \frac{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_1)}{\text{logit}(\tilde{\theta}) - \text{logit}(\rho_2)}$$

ρ_1 thus can be derived as:

$$\rho_1 = \frac{\exp\left[\text{logit}(\tilde{\theta}) + \left(\text{logit}(\rho_2) - \text{logit}(\tilde{\theta})\right) \cdot (\gamma_{\max} + k)\right]}{1 + \exp\left[\text{logit}(\tilde{\theta}) + \left(\text{logit}(\rho_2) - \text{logit}(\tilde{\theta})\right) \cdot (\gamma_{\max} + k)\right]}$$

The initial guesses for parameters $(\rho_1, \rho_2, \gamma_{\max})$ and other pre-specified model properties for different scenarios are listed below.

The first scenario assumes that covariate has impact on toxicity response:

Table 4.3 Model setup for MTD estimation with covariate effect

Intercept	Slope	Target NETS (%)	MTD at max. Cov.	Proportion of TNETS for min. Cov.	Proportion of TNETS for max. Cov.
$\frac{\text{logit}(\tilde{\theta}) - \alpha}{\beta}$	$k = \frac{\delta}{\beta}$	$\tilde{\theta}$	$\gamma_{\max}$	ρ_1	ρ_2
0.5879	0.1879	47.60	0.40	0.12	0.20
0.6875	0.2875	47.60	0.40	0.09	0.20
0.7995	0.1995	47.60	0.60	0.14	0.20
0.8007	0.2007	47.60	0.60	0.25	0.30
0.8883	0.2883	47.60	0.60	0.23	0.30
0.8819	0.2818	47.60	0.60	0.12	0.20

While for the second scenario we assume there is no covariate effect in the model of dose-toxicity relationship:

Table 4.4 Model setup for MTD estimation without covariate effect

TNETS (%)	MTD at max. Cov.	MTD at min. Cov.	Proportion of TNETS for min. Cov.	Proportion of TNETS for max. Cov.
$\tilde{\theta}$	$\gamma_{\max}$	$\gamma_{\min}$	ρ_1	ρ_2
47.60	0.40	0.40	0.20	0.20
47.60	0.40	0.40	0.30	0.30
47.60	0.60	0.60	0.20	0.20
47.60	0.60	0.60	0.30	0.30

We use the MCMC sampling method to simulate the Bayesian framework of the three parameters $(\rho_1, \rho_2, \gamma_{\max})$ of interest. The assessment time window was set up to 28 days. Time and events were simulated using an exponential distribution with mean patient's arrival time of 7 days.

4.4 Simulation Results

Simulation study was conducted with a replicate of 1,000 trials. It shows a promising result that our proposed EWOC-NETS-TITE with covariate effect could correctly identify the covariate effect and successfully suggest a range of MTD for patients with various characteristics. Moreover, as the principle concern of phase I clinical trial designs is the safety of the trial, we also compare the probability of overdosing a patient between our model and basic EWOC-NETS-TITE. The comparison confirms that considering covariate effect in the dose-toxicity relationship could effectively control the overdose probability. In this section, we will display our results by different scenarios: assumption of existing covariate effect and assumption of no covariate effect.

4.4.1 Assumption of Existing Covariate Effect

Since both binary and continuous covariates share the same MTD equation $\gamma_z = \frac{\text{logit}(\tilde{\theta})-\alpha}{\beta} - \frac{\delta}{\beta} \cdot z$, we defined the general equation without considering the format of covariate effect (binary or continuous). Table 4.5 displays the simulation result of MTD estimation with the model considering binary covariate effect and the model ignoring binary covariate effect. And table 4.6 illustrates the estimation of MTD by considering/ignoring the continuous covariate effect.

Overall, our proposed model EWOC-NETS-TITE with covariate could effectively identify the patients' covariate impact, and suggests a reasonable coefficient to adjust the MTD. For binary covariate assumption, when the true MTD pair (γ_0, γ_1) is set up with higher values, such as (0.8, 0.6), (0.6, 0.5), and (0.8, 0.2), the estimated covariate effects are very close to the true value with a small fluctuation. The estimated optimal dose is closer to the true value when MTD is also predefined with higher values. The estimated MTDs for these scenarios are very preferable because they are lower than the true values meaning there is lower risk for patients to be overdosed in the trial. The model without covariate effect tends to be more aggressive, which recommends dose levels approximate to upper bound of the true range of MTD.

Table 4.5 Comparison of MTD estimation with/without binary covariate

True Parameters					Estimated MTD Equation with Covariate Effect $c = 0, 1$	Estimated MTD Ignoring Covariate
ρ_0	ρ_1	γ_0	γ_1	k		
0.30	0.36	0.60	0.40	0.20	$\gamma_c = 0.5808 - 0.1619c$	0.6040
0.10	0.14	0.80	0.60	0.12	$\gamma_c = 0.7108 - 0.1006c$	0.7422
0.10	0.12	0.60	0.50	0.06	$\gamma_c = 0.5877 - 0.0714c$	0.6411
0.10	0.28	0.80	0.20	0.48	$\gamma_c = 0.7159 - 0.4020c$	0.4728

Table 4.6 Comparison of MTD estimation using model with /ignoring covariate

True Parameters			True MTD Equation with Covariate Effect	Estimated MTD Equation with Covariate Effect	Estimated MTD Ignoring Covariate
ρ_1	ρ_2	$\gamma_{\max}$			
0.12	0.20	0.40	$\gamma_c = 0.5879 - 0.1879c$	$\gamma_c = 0.6357 - 0.1648c$	0.6205
0.09	0.20	0.40	$\gamma_c = 0.6875 - 0.2875c$	$\gamma_c = 0.7295 - 0.1837c$	0.6515
0.23	0.30	0.40	$\gamma_c = 0.6305 - 0.1298c$	$\gamma_c = 0.6703 - 0.1132c$	0.6153
0.20	0.30	0.40	$\gamma_c = 0.6749 - 0.1671c$	$\gamma_c = 0.7311 - 0.1566c$	0.6388
0.14	0.20	0.60	$\gamma_c = 0.7995 - 0.1995c$	$\gamma_c = 0.8281 - 0.1791c$	0.7532
0.25	0.30	0.60	$\gamma_c = 0.8007 - 0.2007c$	$\gamma_c = 0.8427 - 0.2069c$	0.7739
0.23	0.30	0.60	$\gamma_c = 0.8883 - 0.2883c$	$\gamma_c = 0.8656 - 0.2265c$	0.6905
0.12	0.20	0.60	$\gamma_c = 0.8819 - 0.2818c$	$\gamma_c = 0.9065 - 0.2457c$	0.7111

Table 4.7 shows the performance of our estimation for the parameters (γ_0, γ_1) with respects to bias and MSE, according to binary covariate effect. The estimation of MTD is very accurate, especially for the MTD for the baseline group — γ_0 that assures the safety of the trial.

And table 4.8 shows the accuracy of the estimation for the parameters $(\rho_1, \rho_2, \gamma_{\max})$ in terms of continuous covariate effect. The bias and MSE are higher when a lower or non-significant covariate is used in the model, the same trend also found when the true MTD is more conservative.

Table 4.7 Accuracy of estimation for $(\boldsymbol{\rho_0}, \boldsymbol{\rho_1}, \boldsymbol{\gamma_{\max}})$ with binary covariate

True Parameters		Considering Covariate				Ignoring Covariate	
		Bias		MSE		Bias	MSE
$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}$	$\hat{\gamma}$
0.6000	0.4000	0.0242	0.0623	0.4019	0.2013	0.0040	0.3040
0.8000	0.6000	-0.0892	0.0185	0.4631	0.3625	-0.0578	0.4422
0.6000	0.5000	-0.0123	0.0162	0.3600	0.3000	0.0411	0.3411
0.8000	0.2000	-0.0841	0.1867	0.4471	0.1451	-0.3272	0.2728

Table 4.8 Accuracy of estimation for $(\boldsymbol{\rho_1}, \boldsymbol{\rho_2}, \boldsymbol{\gamma_{\max}})$

True Values			Bias			MSE		
ρ_1	ρ_2	$\gamma_{\max}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\gamma}_{\max}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\gamma}_{\max}$
0.1200	0.2000	0.4000	0.0431	0.0419	0.0709	0.0139	0.0071	0.0271
0.0900	0.2000	0.4000	0.0606	0.0483	0.1458	0.0221	0.0085	0.0388
0.2300	0.3000	0.4000	0.0068	-0.0289	0.1571	0.0001	0.0010	0.0259
0.2000	0.3000	0.4000	0.0381	-0.0181	0.1746	0.0015	0.0005	0.0310
0.1400	0.2000	0.6000	-0.0016	0.0594	0.0490	0.0094	0.0078	0.0040
0.2500	0.3000	0.6000	-0.0123	-0.0163	0.0358	0.0140	0.0091	0.0042
0.2300	0.3000	0.6000	0.0063	-0.0089	0.0391	0.0002	0.0002	0.0034
0.1200	0.2000	0.6000	0.0094	0.0879	0.0608	0.0001	0.0004	0.0035

Table 4.9 through table 4.10 show the probability of patients being overdosed. Both models EWOC-NETS-TITE with/without covariate have well controlled the overdose probability. As all simulated trials are pre-specified with a target NETS 0.476, and no more than this proportion of patients by both methods have been overdosed. The proportions of overdose are very close with/without consideration of covariate in the model, but overall most proportions are lower if take covariate into account.

Table 4.9 Proportion of overdosing with/without continuous covariate in the model

Predefined Parameters $(\rho_1, \rho_2, \gamma_{\max})$	TNETS $\tilde{\theta}$	Prob. of Overdosing with Covariate	Prob. of Overdosing No Covariate
(0.12, 0.20, 0.40)	0.4760	0.3767	0.3867
(0.09, 0.20, 0.40)	0.4760	0.4417	0.3983

(0.25, 0.30, 0.60)	0.4760	0.1132	0.2783
(0.23, 0.30, 0.60)	0.4760	0.1566	0.3200
(0.14, 0.20, 0.60)	0.4760	0.1017	0.1117
(0.12, 0.20, 0.60)	0.4760	0.0117	0.1283
(0.25, 0.30, 0.60)	0.4760	0.0267	0.0100
(0.23, 0.30, 0.60)	0.4760	0.0250	0.0150

Table 4.10 Proportion of overdosing with/without continuous covariate in the model

True Parameters		Considering Covariate		Ignoring Covariate	
γ_0	γ_1	$c = 0$	$c = 1$	$c = 0$	$c = 1$
0.6000	0.4000	0.3167	0.1567	0.4967	0.2315
0.8000	0.6000	0.1556	0.1111	0.3833	0.1042
0.6000	0.5000	0.2053	0.1334	0.3303	0.1286
0.8000	0.2000	0.1028	0.0093	0.4280	0.0600

4.4.2 Assumption of No Covariate Effect

Using the result from the assumption that no covariate effect is truly applied, we could answer the following questions: is the estimated covariate coefficient small enough by our model with consideration of covariate effect so that the impact of covariate could be ignored; and as well as what is the proportion of overdose in the model. Table 4.11 displays the result of MTD estimation with model considering binary covariate effect and model ignoring binary covariate effect. And table 4.12 illustrates the estimation of MTD by considering/ignoring the continuous covariate effect.

Overall, our proposed model EWOC-NETS-TITE with covariate could effectively identify the patients' covariate impact, and suggests a reasonable coefficient to adjust the MTD. For binary covariate assumption, when the true MTD is set up above 0.6, the estimated MTD very close to the true value. And the results are very satisfactory by both binary and continuous covariate model. The coefficient estimations are very low, which could correctly reflect the reality. The model without covariate effect tends to be more aggressive, which recommends dose levels approximate to upper bound of the true range of MTD.

Table 4.11Comparison of MTD estimation no binary covariate assumption

True Parameters		Estimated MTD Equation with Covariate Effect	Estimated MTD Ignoring Covariate
ρ_0	γ		
0.10	0.20	$\gamma_c = 0.3057$ - 0.0167c	0.3700
0.10	0.40	$\gamma_c = 0.4704$ - 0.0427c	0.5332
0.10	0.60	$\gamma_c = 0.6438$ - 0.0124c	0.6679
0.10	0.80	$\gamma_c = 0.7627$ - 0.0026c	0.7380

Table 4.12 Comparison of MTD estimation no continuous covariate assumption

True Parameters		Estimated MTD Equation with Covariate Effect	Estimated MTD Ignoring Covariate
ρ_0	γ		
0.20	0.40	$\gamma_c = 0.5592$ - 0.0399c	0.5344
0.30	0.40	$\gamma_c = 0.5874$ - 0.0798c	0.5575
0.20	0.60	$\gamma_c = 0.7463$ - 0.1199c	0.6871
0.30	0.60	$\gamma_c = 0.7206$ - 0.1194c	0.7739

Table 4.13 shows the performance of our estimation for the parameters $(\rho_0, \rho_1, \gamma_0)$ with respects to bias and MSE, and table 4.14 shows the estimation accuracy according to continuous covariate. The bias and MSE for model considering covariate is lower than those in the model without considering covariate effect.

Table 4.13Accuracy of estimation for parameters no binary covariate assumption

True Parameters	Considering Covariate				Ignoring Covariate	
	Bias		MSE		Bias	MSE
γ	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}$	$\hat{\gamma}$
0.20	0.1050	0.0638	0.0429	0.0538	0.1700	0.0294
0.40	0.0798	0.0393	0.0331	0.1254	0.1332	0.0182
0.60	0.0385	0.0250	0.0048	0.0933	0.0795	0.0068
0.80	-0.0173	-0.0204	0.0025	0.0105	-0.0220	0.0015

Table 4.14Accuracy of estimation for parameters no continuous covariate assumption

True Values		Bias			MSE		
ρ_0	$\gamma_{\max}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\gamma}_{\max}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\gamma}_{\max}$
0.2000	0.4000	0.0397	0.0537	0.1193	0.0016	0.0032	0.0152
0.3000	0.4000	-0.0618	-0.0345	0.1076	0.0038	0.0015	0.0122
0.2000	0.6000	0.0385	0.0710	0.0264	0.0015	0.0052	0.0011
0.3000	0.6000	-0.0628	-0.0292	0.0012	0.0039	0.0011	0.0014

Table 4.15 through table 4.16 show that the probability of patients being overdosed. Both models EWOC-NETS-TITE with/without covariate have well controlled the overdose

probability. As all simulated trials are pre-specified with a target NETS 0.476, and no more than this proportion of patients by both methods have been overdosed. The proportions of overdose are very close with/without consideration of covariate in the model, but overall most proportions are lower if take covariate into account. The result of overdose probability control may due to the contribution of EWOC idea.

Table 4.15Proportion of overdosing no binary covariate assumption

Predefined Parameters (ρ_0, γ)	TNETS $\tilde{\theta}$	Prob. of Overdosing with Covariate	Prob. of Overdosing No Covariate
(0.10, 0.20)	0.4760	0.4333	0.2400
(0.10, 0.40)	0.4760	0.2133	0.1067
(0.10, 0.60)	0.4760	0.0133	0.0300
(0.10, 0.80)	0.4760	0.0001	0.0020

Table 4.16 Proportion of overdosing no continuous covariate assumption

Predefined Parameters (ρ_0, γ)	TNETS $\tilde{\theta}$	Prob. of Overdosing with Covariate	Prob. of Overdosing No Covariate
(0.20, 0.40)	0.4760	0.5700	0.5217
(0.30, 0.40)	0.4760	0.4667	0.4783
(0.20, 0.60)	0.4760	0.1217	0.0450
(0.30, 0.60)	0.4760	0.0433	0.0150

4.5 Conclusion and Discussion

The advantage of our proposed model is quite explicit that it suggests a range of dose (in binary covariate cases, it suggests different dose for each subgroup of patients) for according to patient's characteristics. The flexibility is beneficial for patients who are sensitive to toxicity and a conservative level of dose could protect them from the risk of overdosing; it is also beneficial for patients who are able to tolerate a higher dose because the treatment effect could be optimized.

The principle guidance for a cancer phase I trial is to preserve the safety as well as to estimate a MTD. Now we give this goal another condition: patient's covariate effect. Using the EWOC-NETS-TITE with binary covariate model could accurately estimate the MTD for patients

in different groups. And the EWOC-NETS-TITE with continuous covariate model, which is more general, could help clinicians allocate dosage with more flexibility and more accuracy.

The other advantage of the proposed model is that it prevents more patients from being overdosed. The adjustment according to patients' heterogeneity provides a broader range that allows patients with different health conditions could be treated to their optimal benefit. This conclusion could be easily seen from the probability of overdosing a patient. We compare different scenarios; one assumes that the patients' heterogeneity has true impact on patients' dose-toxicity reaction. Then under this condition, the model considering covariate effect is definitely superior to the model without the covariate. And the simulation results shows that the impact could be accurately identified by the EWOC-NETS-TITE with covariate design. When the true MTD is relatively low, such 0.2, or 0.4, our estimation tends to be higher than the true value. The possible reason is when the true MTD is too low, it is really hard to apply treatment to patients without any escalation in the trial, especially when the initial dose is very close to the true MTD. In other conditions, our estimation is not only accurate, but also slightly conservative to ensure the patients' safety. We also compare the results using models with or without considering the covariate effect when the assumption is patients' heterogeneities have no impact on the dose-toxicity reaction. We got very satisfactory results from the simulation study. The model returns very small coefficient that will cost very little loss to the estimation of MTD. We also compare our results to a published method — EWOC with binary covariate (Tighiouart, et al., 2012). Our proposed design is more precise, with smaller bias and MSE which dues to the advantages in the method of NETS. In the further research direction, we would like to propose more complex model considering multiple patients' characteristics with binary and continuous covariate incorporated.

5 NOVEL TWO-STAGE PHASE II DESIGN

A phase II trial is an expeditious and low cost trial with the primary goal of screening potentially effective agents prior to confirmatory phase III trial. However, the success rate of phase III oncology trials remains very low despite the success demonstrated in the preceding phase II trials. This discordance is mainly due to the different endpoints used in phase II (tumor response) and phase III (survival) trials. While a robust disease response is expected to translate into survival improvement, this is not guaranteed. Moreover, tumor response can be determined quickly whereas survival estimate requires a long period of follow up. We propose a novel two-stage screening design for phase II trials in this dissertation whereby percent of tumor size change endpoint is used as an initial screening to select potentially effective agents within a short time interval followed by a second screening stage where progression free survival is estimated to confirm the efficacy of agents. This design can improve trial efficiency and reduce cost by early stopping the evaluation of an ineffective agent based on low percent of tumor size change. The second survival endpoint screening will substantially increase the success rate of follow-up phase III trial by using the similar outcomes. We conducted simulation studies to investigate the

underlying statistical considerations to optimize the significant levels of the two screening stages in the design.

5.1 Introduction

The main purpose of a cancer phase II clinical trial is to evaluate the preliminary anticancer efficacy of new drug or treatment administered at the MTD and schedule estimated in proceeding phase I clinical trial. The successful drug or treatment that screened out from the phase II trial will be further confirmed for effectiveness and long term adverse effect in a large phase III trial. As a screening trial of subsequent phase III trial, phase II trial usually enrolls fewer than a hundred participants. Some randomized phase II trial or phase II/III trial may enrolled up to hundreds (Rubinstein, et al., 2005).

The tumor response rate has been widely adopted as the primary endpoint in phase II clinical trial, assuming that higher response rate in phase II trials associates with longer survival time which is the gold standard endpoint of the following phase III trial. In conventional phase II trials, tumor shrinkage between the baseline and the measurement after treatment is measured and categorized into four categories (complete response (CR), partial response (PR), stable disease (SD), and progression disease (PD) according to the Response Evaluation Criteria in Solid Tumors (RECIST 1.1) (Eisenhauer, et al., 2009) or World Health Organization (WHO) criteria (WHO, 1979). The proportion of patients that have objective responses who achieve CR or PR is defined as response rate. Later in this chapter, we will compare the performance of our design with a popularly used design — Simon's two-stage design. Simon's two-stage design (Simon, 1989) uses WHO criteria for response rate calculation; to keep consistency, we adopt WHO criteria for our simulation study as well.

Today, cancer is the still the leading cause of death worldwide after 40 years of unremitting study, accounting about 13% of all death in 2008 (WHO, 2013). Any new anticancer drug and treatment must be proved to be significantly effective in the confirmative phase III clinical trial before they can be administered to patients. At present, the failure rate of phase III in oncology trials remains very high (e.g. 50-60%) despite the success demonstrated in the preceding phase II trials (An MW, 2011). The high failure rate is mainly due to the different endpoints used in phase II (tumor response) and phase III (survival) trials. While a robust tumor response is expected to translate into survival improvement, this is not guaranteed (An MW, 2011). Some study has confirmed significant relationship of response rate with progression free survival and overall survival (10% and 11.4% response rate increments correspond to 1 month increase in PFS and OS, respectively) (Louvet, et al., 2001).

Apprehensions about adopting response rate as a primary endpoint are recently well discussed. First, the simplicity achieved by creating only two response groups via WHO or RECIST criteria has a cost: categorize of continuous data may be consequent with a loss of information (Pivot, et al., 2009). More fundamentally, the ultimate goal of a new drug under development is to prolong survival rather than to raise response rate. These concerns prompted us to propose a new two-stage phase II design, which evaluates percent in tumor size changes as a continuous endpoint in Stage I and estimates PFS in stage II. Our approach that includes a screening stage of survival time in order to improve phase III success rate is beneficial to both pharmaceutical companies and patients.

Previously researches on tumor size changes as a continuous variable have been proposed to evaluate antitumor activity (Lavin, 1981) (Wang, et al., 2009). We adopt the idea by Wang *et al.*, to model the tumor size, but go one further step by modifying it assuming treatment effect

also prevent tumor cells' self-progression. And a simple *t* test is applied to compare tumor size data in different treatment groups in stage I. Recently many researches in oncology confirm PFS as the best estimate of overall survival (OS) (Buyse, et al., 2000). Although OS remains regulatory gold standard and is more reliable in classifying event status; PFS has the advantage of short median survival and more informative within the protocol (Halabi, et al., 2009) (Yothers, 2007). PFS is a more sensitive indicator of treatment effect and is adopted in stage II screening in our design (Buyse, et al., 2007) (Buyse, et al., 2000).

Before we propose the new two-stage design, a brief review of Simon's two-stage design and criteria of tumor response is made. Table 5.1 shows the current WHO criteria and RECIST criteria for tumor responses. As shown below, 50% of tumor shrinkage will be used as a standard for objective response.

Table 5.1WHO and RECIST criteria for tumor response

Term	RECIST Criteria	WHO Criteria
Target lesions	Measurable lesions to a maximum of five (two per organ)	All measurable lesions
Type of measurement	Unidimensional	Bidimensional
Tumor burden assessment	Sum of greatest diameter of target lesions	Sum of products of maximum perpendicular diameters
Response		
Complete response (CR)	Disappearance of all known lesion(s); confirmed at four weeks.	Disappearance of all known lesion(s); confirmed at four weeks.
Partial response (PR)	At least 30% decrease; confirmed at four weeks.	At least 50% decrease; confirmed at four weeks.
Stable disease (SD)	Neither PR nor PD criteria met	Neither PR nor PD criteria met
Progressive disease (PD)	20% increase; no CR, PR, or SD documented before increased disease, or new lesion(s)	25% increase; no CR, PR, or SD documented before increased disease, or new lesion(s)

The hypotheses of a Simon's design are similar as the general hypotheses of phase II clinical trials that we described in chapter 2. The hypotheses can be stated as:

$$H_0: p \leq p_0 \text{ vs. } H_1: p \geq p_1$$

p_0 is the largest response proportion which if true, clearly implies that the treatment does not warrant further study. It is sometimes called the response rate of a poor treatment. p_1 is the smallest response proportion which, if true, indicates the treatment does warrant further study. It is sometimes called the response rate of a good treatment. α is defined as the probability of rejecting the null hypothesis when it is true (type I error), and β denotes the probability of rejecting the alternative hypothesis when it is true (type II error). Then the Simon's two-stage design scheme is as follow: the design can be represented by four numbers — n_1, n, r_1, r. n_1 is the sample size in the first stage, while r_1 is the critical value in the first stage; n is the combined sample size for both stages, and r is the critical value in the combined sample. If r or fewer of n patients respond, the new drug or treatment is rejected.

The expected sample size of Simon's two-stage design is

$$E(n_E) = n_1 + (1 - PET)(n - n_1)$$

where PET is the probability of early termination of the study. Probability of rejecting the tested drug with true response proportion p can be calculated using binomial distribution with

$$\Pr(\text{reject}|p, n_1, r_1, r, n) = B(r_1|p, n_1) + \sum_{x=r1+1}^{\min(N_1, r)} b(x|p, n_1) B(r - x|p, n - n_1)$$

where $b(x|p,n) = \frac{n!}{x!(n-x)!} p^x (1-p)^x$, and $B(x|p,n) = \sum_{r=0}^{x} b(r|p,n)$.

5.2 Two-Stage Double Screening Phase II Design

5.2.1 Design Scheme

Although many designs have been proposed for phase II clinical trial, the success rates have changed little. The success rate of phase III oncology trials remains very low despite the success demonstrated in the preceding Phase II trials. This discordance is mainly due to the different endpoints used in Phase II (tumor response) and III (survival) trials. While a robust disease response is expected to translate into survival improvement, this is NOT guaranteed. Moreover, disease response can be determined quickly whereas survival estimation requires a long period of follow up. We propose a novel two-stage screening design for phase II trials whereby percent of tumor size change endpoint is used as an initial screening to select potentially effective agents within a short time interval followed by a second screening stage where progression free survival is estimated to confirm the efficacy of agents. Figure 5-1 gives a schematic of our proposed design, which we will describe in this section. This design can improve trial efficiency and reduce cost by early stopping the evaluation of an ineffective agent based on low percent of tumor size change. The second survival endpoint screening will substantially increase the success rate of follow-up Phase III trial by using the similar outcomes.

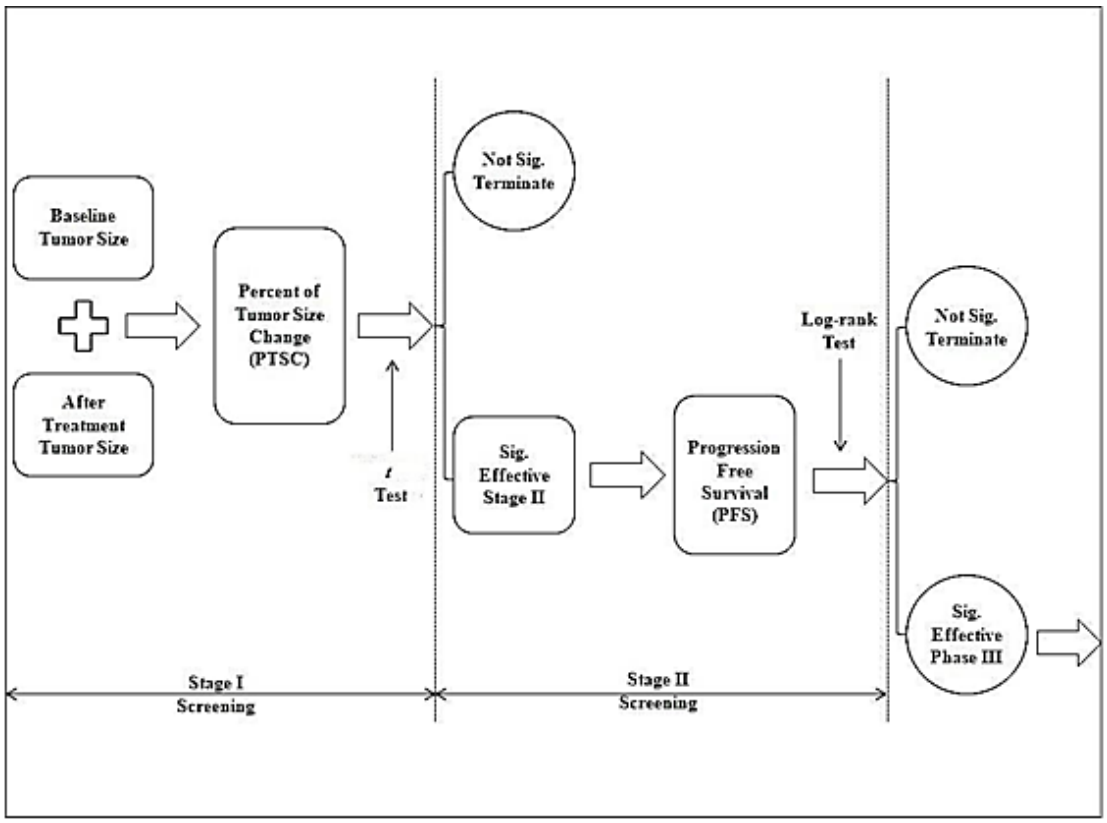

Figure 5-1Two-stage double screening design

The two-stage double screening design can be described as follow: patients who meet the trial "inclusion criteria" with baseline tumor size measurement are given the targeted new treatment. After a predefined period of time, we calculate percent of tumor size change (PTSC) according to both baseline and current tumor size. Stage I screening is taken to the experimental group with new treatment, and to the control group or historical data. A decision will be made to early terminate the trial, if there is no significant result found. Otherwise, the experiment continues to second stage, in which we will investigate difference of PFS. Similarly, the trial will be terminated when no significant difference found, or we will proceed to subsequent phase III trial if the result is significant.

5.2.2 Advantages of Proposed Phase II Design

There are 3 major advantages in the proposed design: 1) The continuous tumor size change percentage is used instead of the categorical tumor response, which will fully utilize tumor change information; 2) The PFS is used as the endpoint in the second stage of the design which could better predict the success of follow up phase III trials; 3) A screening is used in the middle between first stage and second stage. Therefore, the power and length of the trial has been substantially improved, so does the trial efficiency.

5.2.3 Model of Changes in Tumor Sizes

To model the changes in tumor sizes, we adopt an exponential model as a basic distribution for tumor growth at first, and then modify it to accommodate the clinical meanings. The reason to choose exponential model is in clinical studies, it is believed an exponential model is very close to the biological foundation of how cells growth (Skipper, et al., 1982). Then the tumor size is formulated as

$$\hat{V}_i(t) = V_i(0) \cdot \exp(\rho_i \cdot t)$$

where $V_i(t)$ is the i^{th} patient's tumor size at time point t, $V_i(0)$ is baseline tumor size at the starting point of the patient, and ρ_i is the tumor growth rate. The exponential growth pattern is based on the assumption that no death or treatment intervention, and is considered to be appropriate (Sachs, et al., 2001). To simulate data with external agent, Wang *et al.* proposed a mixed exponential-decay and linear-growth model (Wang, et al., 2009)

$$\hat{V}_i(t) = V_i(0) \cdot \exp(-\lambda_i t) + \rho_i \cdot t$$

The above model includes treatment effect by an exponential tumor shrinkage rate λ and a linear tumor progression effect with rate ρ. However, considering that a patient's tumor progression could be inhibited when a treatment intervention is got involved in. We thus

modified the above equation to simulate our data with the considertation that the treatment effect applies to tumor's self-progression as well.

$$\hat{V}_i(t) = (V_i(0) + \rho_i \cdot t) \cdot \exp(-\lambda_i t)$$

Random variability is attributed to two sources: interpatient variability and residual variability. The residual variability captures error caused by model misspecification and/or in tumor measurements. The population is assumed to follow a log-normal distribution.

$$V_i(0) = \mathcal{M}_V \cdot \exp(\eta_i)$$

In this equation, $\mathcal{M}_V$ is the population median baseline tumor size and η_i is the difference between the individual and population median baseline values on a log scale that is assumed to follow a normal distribution with a mean of zero and variance of ω_V^2. The individual parameters of tumor self-progression ρ_i and treatment effect λ_i are also described using similar equations. An exponential error model is used for residual variability.

$$V_i(t) = \hat{V}_i(t) \cdot \exp(\varepsilon_i)$$

As denoted above $\hat{V}_i(t)$ is the observed tumor size at time t for the i^{th} individual, while $V_i(t)$ is the expected tumor size at time t for the i^{th} individual, and ε_i is the difference between the observed and expected values on a log scale. ε_i is assumed to follow a normal distribution with a mean of zero and variance of σ_ε^2.

Therefore, the continuous tumor size modeled with the equation would result in the tumor size asymptotically reducing toward zero. In our model, λ — the rate for tumor shrinkage and ρ — the progression rate are both restricted to be non-negative. Considering tumor growth kinetics, inidividual patient's tumor size is generated from exponential distribution (Friberg, et al., 1997). PTSC is calculated as:

$$p_i(t) = \frac{(V_i(t) - V_i(0))}{V_i(0)} \times 100$$

5.3 Simulation Study

A simulation study is conducted to assess the performance of our proposed two-stage phase II design. There are two major parts in the simulation study. First part of our simulation study is used to compare the model performance and characteristics with a famous and most commonly utilized phase II design — Simon's two-stage design; the second part of our simulation study is used to compare the results with conventional test for stage I screening, evaluate and suggest a cut-off point for the stage I screening. A replicate of 10,000 trials was done for both part of the simulation. Response rates were generated according to WHO criteria (Table 5.1), percent changes in tumor sizes were generated with respect to tumor size model describe above in section 5.2.3.

5.3.1 Simulation Plan

For the first part of our simulation, we conducted a simulation study to compare the model performance with Simon's two-stage design. The motivation of this dissertation research is to seek for a design that could effectively improve the success rate of subsequent cancer phase III trials. Hence, one of our most important missions is to test whether this proposed design raise the follow up phase III trials' success rates.

With this purpose, we have done the following simulation:

(1) Comparison of success rates in follow up phase III trials. In order to estimate the success rate in subsequent phase III trials, we assign each subject in a single trial with a true overall survival. The OS consists of two parts of survival time: progression-free survival time and post-progression survival time. The post-progression survival time is universal, follows a general exponential distribution for all patients. Progression-free survival time, on the other

hand, varies from patient to patient, although for the general population we assume the baseline progression-free survival time is a random variable from an exponential distribution. Whereas we consider a patient's progression-free survival time is also affected by the percent change in tumor size with treatment intervention. We connect the patient final progression-free survival time with PTSC by using a prespecified variable — marginal increment progression-free survival time. For example, 1% tumor shrinkage will result in, for example 0.05 day increment in the mean of the exponential distribution of the progression-free survival time. Figure 4-2 demonstrates this process.

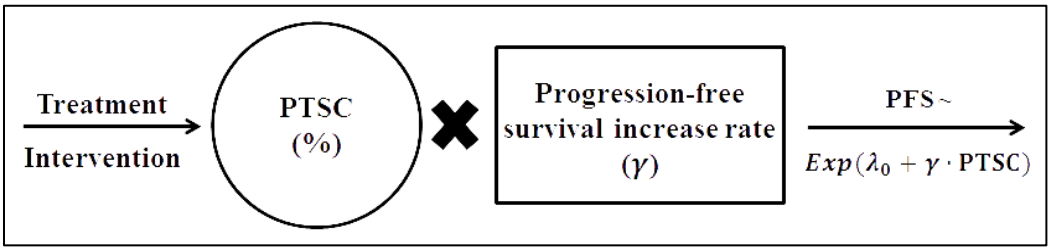

Figure 5-2 Progression-free survival time generation

Therefore, the overall survival time in subsequent phase III trial is the sum of progression-free survival time and post-progression survival time, with percent censored being 10%. Then we use overall survival as a golden standard to estimate the true outcome in the follow up phase III trial. Using a contingency table, we thus can specify both designs performance by inputting the numbers of true positive, false positive, true negative and false negative counted in the simulated data, then the success rate can be estimated.

(2) A comparison between the model efficacies has also been made. Because of the complexity of our tumor size model. It is very hard to derive the sample size requirement of the design. Hence, to contrast the efficacy of both Simon's and our two-stage design, a comparison between model sensitivity, specificity, and accuracy is made instead by using Simon's two-stage design sample size for each stage in both designs.

The second part of our simulation study examines the performance of the proposed two-stage design over the conventional test (χ^2 test) on response rate. In this part of simulation, we construct a pseudo two-stage design which uses conventional χ^2 test on tumor responses in first screening stage. PFS is the primary endpoint in second stage of both designs in comparison. We tested different scenarios of tumor responses, and examines how these scenarios affects the *P*-values of the design tests. Moreover, we explore what is the impact of a selection of α level in first screening stage on the final outcome after second screening done.

5.3.2 Simulation Setup

As mentioned above, when comparing the model performance with Simon's two-stage design, sample sizes used in the simulation study were calculated using Simon's optimal two-stage design. The sample sizes in each stage are shown in Table 5.2 for different scenarios. Notations are adopted in previous section 5.1, but made some modification to make simulation setup more straightforward.

Table 5.2 Sample size setup for simulation study

p_0	p_1	n_1	n_2	r_1	r
0.10	0.20	24	41	2	9
0.15	0.25	29	55	4	16
0.20	0.30	39	60	8	24
0.25	0.35	43	69	11	33
0.05	0.20	9	15	0	2
0.10	0.25	13	21	1	5
0.15	0.30	19	20	3	8
0.20	0.35	13	33	2	12
0.05	0.25	6	17	0	2
0.10	0.30	7	11	0	3
0.15	0.35	9	14	1	5
0.20	0.40	12	13	2	7

Notations used in Table 5.2 include n_1, n_2, r_1, r. n_1 is the sample size in the first stage, while r_1 is the critical value in the first stage; n_2 is the sample size for second stage, and r is the

critical value in the combined sample. If r or fewer of n patients respond, the new drug or treatment is rejected.

In order to obtain the simulated data with a desired proportion in tumor responses (for example, set up a desired true response rate as 10%, then 10% of patients in all 10,000 trials should have at least 50% tumor shrinkages), we use the proposed model for tumor size change in section 5.2.3. We assume the general population shares the same baseline tumor size distribution, self-progression rate. When comparing the model performance, the random values of tumor sizes were generated from the tumor size model with the same parameter values except treatment effect factor λ. Table 5.3 illustrates the parameters used to generate the tumor sizes in the simulation study.

Table 5.3 Parameter estimate for tumor size model

p	$\mathcal{M}_V$	$\mathcal{M}_\rho$	$\mathcal{M}_\lambda$	ω_V	ω_ρ	ω_λ	σ_ε
0.05	9.60	1.20	0.02782	0.68	0.67	0.53	0.13
0.10	9.60	1.20	0.03313	0.68	0.67	0.53	0.13
0.15	9.60	1.20	0.03734	0.68	0.67	0.53	0.13
0.20	9.60	1.20	0.04104	0.68	0.67	0.53	0.13
0.25	9.60	1.20	0.04450	0.68	0.67	0.53	0.13
0.30	9.60	1.20	0.04787	0.68	0.67	0.53	0.13
0.35	9.60	1.20	0.05125	0.68	0.67	0.53	0.13
0.40	9.60	1.20	0.05450	0.68	0.67	0.53	0.13

5.4 Simulation Result

In the first part of this section, we will demonstrate the performance of our two-stage design by comparing one of the most commonly used phase II trial designs — Simon's two-stage design. While the later part of this section is the simulation result of our proposed design contrast to conventional phase II clinical trial design.

5.4.1 Success Rate in Follow Up Phase III Trials

As we know, the null and alternative hypotheses of Simon's two-stage design are used to compare the true response rate with an uninteresting and a target response rate, which is described as follow:

$$H_0: p \leq p_0 \text{ vs. } H_1: p \geq p_1$$

In order to evaluate the performance of our proposed design, we need to construct the equivalent hypotheses. Considering the endpoints selected in the first and second screening stages are PTSC and PFS, respectively. We tested the hypotheses as follow:

$$H_0: \mu_p \leq \mu_{p_0} \text{ vs. } H_1: S_p(t) \geq S_{p_1}(t)$$

In Simon's design the first screening stage determines whether the new drug or treatment is better or have a higher response rate using a prespecified criteria (p_0); and a condition should be emphasized here is: Simon's two-stage design is a single-arm design for phase II clinical trials. Therefore, when we did the simulation study in stage I, it should meet the above two requirements. In other words, there was no control group, only one treatment group in the experiment. We use an example here to break down the steps:

(1) For Simon's design, we choose a series of three response rates: uninteresting response rate p_0, target response rate p_1 and true response rate p. Use the combination of the first two rates; we could get sample size needed for the simulation setup.

(2) Use the sample size and parameters estimated table to generate 10,000 trials. Since the true response rate is the mean response rate based on large quantity of experiments, the response rate in the generated dataset is very close to p. Then the mean percent change in tumor size could be estimated. Repeat this step for about 10 times to confirm the stabilization of the μ_p.

(3) Use large enough sample size (e.g. 50 patients, or 100 patients) and the parameters estimated table to generate 10,000 trials of response rate p_1. Solve for the mean PTSC by adopting the method described in step 2.

Therefore, the first screening stage null hypothesis can be tested using a one sample t test. Similarly, we use the same steps to solve survival time in the second stage. However, the problematic issue we encountered in stage II is we fail to find a method to do one sample Log-rank test. So we use a pseudo control group with response rate p_1 in the second screening stage. This problem will leave us as one the further research area. Yang & Zhao has proposed to use weighted log rank test for testing treatment effect which directs us a very promising alternative solution to it (Yang, et al., 2007).

Table 5.4 illustrates subsequent phase III trials' estimated success rates followed by the above hypotheses.

Table 5.4Comparison of success rates of phase III trials Simon's and proposed two-stage design

P_0	P_1	P	Simon's Two-Stage Design	Proposed Two-Stage Design
0.10	0.20	0.15	16.96%	68.57%
0.10	0.20	0.25	7.16%	24.14%
0.20	0.30	0.25	5.55%	29.63%
0.25	0.35	0.30	5.51%	83.33%
0.05	0.20	0.25	15.86%	67.59%
0.10	0.25	0.20	2.56%	56.25%
0.15	0.30	0.25	3.12%	54.55%
0.20	0.35	0.30	2.69%	51.72%
0.05	0.25	0.20	4.45%	45.28%
0.10	0.30	0.35	16.18%	74.81%
0.10	0.30	0.25	16.34%	73.68%
0.15	0.35	0.25	2.99%	46.67%
0.15	0.35	0.30	3.29%	67.74%
0.15	0.35	0.40	2.83%	51.43%
0.20	0.40	0.35	2.38%	46.67%
0.20	0.40	0.45	2.60%	66.67%

The results shown in the above table illustrates the main advantage of our phase II two-stage design. Success rates in follow up phase III trials estimated by our proposed two-stage design vary from 24.14% to as high as 83.33%. Majority of the success rates calculated are above 50%, which performs better than all general methods for cancer phase II trials (50 – 60% of all phase III trials under investigation fail). On the other hand, success rates of Simon's two-stage ranges from 2.38% to 16.96%, which are much lower than the results of the proposed new two-stage design.

There are several reasons that may lead to the difference between success rates in the two methods. We analyze the most possible ones as follow:

(1) First and foremost, improvement in success rates using the proposed two-stage phase II design is that the assumption of overall survival time. We generated the overall survival in phase III trials based on the assumption that patients' overall survival time is the summation of progression-free survival time and post-progression survival time. The first screening stage results have not been affected very much by this assumption. Because for both designs' first stage, the interest is on tumor size changes, although our design used continuous percent of changes and Simon's design used categorical tumor responses as primary endpoint. It is the stage II screening determined the comparison result eventually. Using PFS as endpoint made our proposed design superior than Simon's design without doubt, although we assume a connection between tumor size change and progression-free survival as well. But the later assumption is not as direct as the former one.

(2) The other reason underlying is the motivation of the two designs. The purpose of Simon's two-stage design is to evaluate the new drug or treatment efficacy while by using the

optimal sample size. In the other word, during the process of pursuing an optimal or minimal sample size, it may sacrifice the capability to detect or filter out the confounding results.

Except the above analysis of the significant difference between these two designs. We also notice that the success rates that estimated from our simulation study have a relatively large range. Therefore, we go through our simulation plan and model set up. The potential reason is the changes in tumor sizes vary very much in distributions when different parameter estimates were implemented. For example, when the response rate are low meaning that observations spread all over the range from -100 to $+\infty$, then both first and second screening stage would probably recognize the trial as promising as they are associated. However, when testing overall survival in follow up phase III trials, the result is opposite. Overall survival consists of two parts — progression-free survival time and post-progression survival time. Since response rate are low, the improvement in progression-free sometimes can be neglected. But post-progression survival time is generated based on common population. Hence, the success rate decreases below 50%.

5.4.2 Model Sensitivity

In clinical trials, an important element to evaluate a testing method for treatment effect is model sensitivity. Specifically, we could use sensitivity, specificity, and accuracy as indicators to compare two designs' performance. On the other hand, statistical power and sample size are also very important components to assess a design. However, since the model we used to generate tumor sizes for patients is complicated at some extend, it is hard to derive the sample size and calculate the statistical power directly. Therefore, all the simulations had been done for both designs (Simon's two-stage design and our proposed two-stage design) uses the same sample size according to different scenarios of Simon's two-stage design. The logic here is: if same sample size is used in both designs, then the one with better statistical measures of the

performance (sensitivity, specificity, and accuracy) is the more efficient and powerful design. It should also have the advantage of using fewer samples to gain the same statistical power.

The steps of how we simulate the clinical trials were described in section 5.5.1, and we adopt the same procedure to generate and compare our data. Table 5.5 illustrates the simulated results.

Table 5.5 Comparison of sensitivity, specificity, and accuracy in Simon's two-stage design and our proposed two-stage design

			Simon's two-stage design			Novel two-stage design		
p_0	p_1	p	sensitivity	specificity	Accuracy	sensitivity	specificity	Accuracy
0.10	0.20	0.15	87.42%	16.96%	27.60%	47.68%	96.11%	88.80%
0.10	0.20	0.25	97.73%	3.64%	20.20%	64.20%	94.54%	89.20%
0.15	0.25	0.20	97.43%	13.27%	26.40%	64.10%	95.60%	90.10%
0.20	0.30	0.25	30.00%	46.63%	45.80%	24.00%	98.52%	94.80%
0.20	0.30	0.35	92.11%	1.56%	5.00%	51.35%	98.13%	96.40%
0.05	0.20	0.25	95.54%	5.58%	5.60%	62.42%	94.42%	94.80%
0.10	0.25	0.20	43.24%	36.66%	10.10%	24.32%	99.27%	91.70%
0.15	0.30	0.25	56.41%	32.57%	7.70%	30.77%	98.02%	93.00%
0.20	0.35	0.30	96.30%	3.39%	33.50%	55.56%	98.56%	95.40%
0.05	0.25	0.20	71.70%	13.83%	7.10%	45.28%	96.94%	94.30%
0.10	0.30	0.35	99.38%	0.48%	7.30%	62.35%	95.94%	92.20%
0.10	0.30	0.25	92.86%	3.97%	10.70%	41.67%	97.00%	92.40%
0.15	0.35	0.25	61.90%	28.71%	30.10%	35.71%	98.64%	96.00%
0.15	0.35	0.30	92.00%	28.84%	7.10%	42.00%	98.95%	95.10%
0.15	0.35	0.40	90.00%	4.54%	27.60%	60.00%	98.25%	88.80%
0.20	0.40	0.35	47.06%	32.09%	20.20%	20.59%	99.17%	89.20%
0.20	0.40	0.45	92.59%	3.80%	26.40%	66.67%	99.08%	90.10%

Our proposed two-stage design has shown a series of very promising specificity and accuracy using simulation study. The estimated specificities, which are the proportion of the true negatives that correctly identified by our proposed design are all above 90%. This specificity suggests how good our design is at identifying the normal (negative) condition. A high specificity indicates that our proposed two-stage design could properly identify the invalid drugs or treatments, and prevent a further large phase III trials from huge spends and time on research of inefficient drugs or treatments. The accuracy which represents the proportion of true results,

either true positive or true negative measure the degree of veracity of a design on certain scenario. Table 5.5 shows that the proportions of true results that were successful detected by our proposed design were as high as 96.40%, and majority of the accuracies calculated were around 90%. However, the sensitivity which represents the probability of the design identifies candidate drugs which are effective is relative low comparing to Simon's two-stage design. Most sensitivity was around 40-50%, but some of them can be as low as 20%. While Simon's two-stage tends to have the opposite results.

Comparing results from both designs, Simon's two-stage tends to allow more possible candidate drugs take part into the follow up phase III trials. The advantage of it is this design gives opportunities to more potential drugs for further investigation. However, these opportunities can also be considered as a waste of resource, because the proportion of false positive results is high, meaning that large numbers of tested new drugs entering phase III would not have a confirmatory results. As phase III trials are the most expensive and prolonged trial among the all three. A high false positive rate is definitely should be avoid. Considering both Simon's and our two-stage designs, because of the distinctions between the design purposes, the result of interests are very much different. Simon's two-stage design aims to "pick out" as many potential treatments as possible for further investigation, while our proposed two-stage design aims to "avoid" selecting inefficient agents so that improve the success rates of follow up phase III trials. If balancing all measures, our proposed design is superior to Simon's design. But in further research plan, our goal could be modified from focusing on success rate only to balancing sensitivity and success rate.

5.4.3 Comparison with Conventional Design

A comparison between the two-stage phase II trials has also been done via simulation

study with different measurements of tumor shrinkages (continuous measurement of percent change in tumor size vs. grouped tumor response rate). Figure 5-3 demonstrates success rate in the second stage by various significant level in first stage. It shows the success rate of the method that uses continuous endpoint (PTSC) directly is superior to the method that uses categorical endpoint (response rate) in the first screening stage.

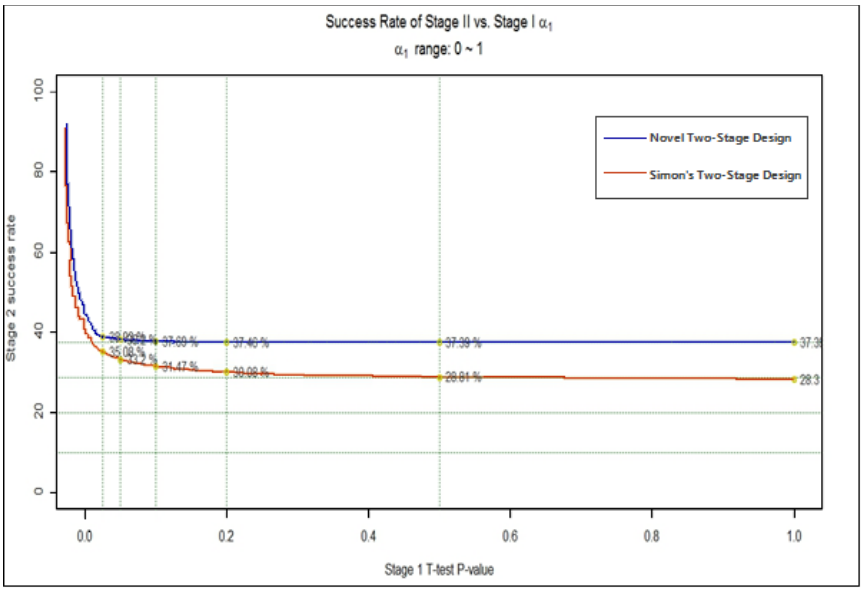

Figure 5-3Success rate in second stage by various significant level in first stage

The half bath-tube shaped curve in Figure 5-3 implies a very sharply decreasing at first and then very smooth trend for the success rate. From 0 to 0.025 the success rate declines more than 30%, while another 10% decrease happened between the range of 0.025 and 0.2. In contrary, the cumulative decrease can be ignored starting from 0.2 till 1.With this result, we would like to suggest a rigorous significant level for first screening stage (e.g. $\alpha_1 \leq 0.05$). It is the general belief that if we set up loose criteria when filtering the potential treatments, then there is a better chance to find more promising results in later phase III trials. However, this assumption is not true from interpretation of our simulation result. Therefore, a restricted significant level should be set up to improve the success rate of the overall phase II trials.

5.5 Conclusion and Discussion

In this section, we will discuss and conclude our propose two-stage phase II design using the above results. We will discuss the following topic in this section: selection of endpoint, double screening design scheme, and critical values setup.

5.5.1 Percent Changes in Tumor Sizes as Endpoint in First Stage

In clinical studies, categorization is quite commonly adopted by grouping continuous values into ≥ 2 categories (Naggara, et al., 2011), not only limited to cancer phase II trials. The primary reason according to the approach is the need to label patients with an attribute for diagnostic or therapeutic procedures determination (e.g. 'hypertensive', 'obese') (Royston, et al., 2006). However, the disadvantage of grouping continuous variables is obvious. A serious loss of power (Lagakos, 1988) and higher sample size requirements (Wason, et al., 2011) in effectively detecting possible relationships is the cost of simplicity. At least one third, even higher proportion if the predictor is exponentially distributed of the data is discarded while dichotomizing (Lagakos, 1988). Additionally, cutoff point selection of the categories is controversial when concerning about clinical benefit behind it. For example, a patient with 25% tumor shrinks usually has more clinical benefit than one with 10% tumor increases, but both are labeled by WHO criteria (WHO, 1979) as having stable disease. In contrary, a patient is identified as an objective responder with 55% shrinkage may not has much difference with the one with 45% shrinkage, but the latter is not (Karrison, et al., 2007). Therefore, it is not surprising that more and more researchers choose to use tumor size changes directly instead of response rate.From a statistical standpoint, categorizing a continuous tumor change percentage into a categorical tumor response with 4 levels results in a loss of study power by not fully utilizing all available data. Several publications have studied extensively the direct utilization of

continuous tumor shrinkage as the primary endpoint for the measurement of drug efficacy in phase II clinical trials (Wang, et al., 2009) (Karrison, et al., 2007) (Lavin, 1981).

5.5.2 PFS as Surrogate Endpoint of OS in Second Stage

Overall survival is the traditional and the objectively measured endpoint that is adopted to assess new cancer drugs. However, it requires prolonged follow-up and so may not be optimal for a fast assessment of therapeutic advances (Burzykowski, et al., 2004). Moreover, many clinical trials now include sequential therapies, and overall survival as a primary endpoint would not accurately reflect the effect of the investigational drug with multiple lines of treatment (Hotte, et al., 2011). So many researchers proposed surrogate endpoints for overall survival to evaluate the clinical benefits of new drugs in oncology, where PFS is frequently adopted. For example, Gill *et al.* used PFS in clinical trials of metastatic colorectal cancer (Gill, et al., 2011), and Saad *et al.* reviewed PFS as a surrogate endpoint in breast and colorectal cancer treatment (Saad, et al., 2010). So far, the success rate of phase III oncology trials remains very low (e.g. 50-60%) despite the success demonstrated in the preceding phase II trials. The relationship between tumor response/tumor shrinkage percentage and overall survival as the gold standard for drug efficacy has been revisited (An MW, 2011). PFS has the advantage of short follow up time (Yothers, 2007) and has been confirmed as the best estimate of overall survival (Buyse, et al., 2000). We thus choose to use PFS as the primary endpoint in the second stage in our phase II design.

5.5.3 Double Screening

The advantage of bouble screening is to allow a prestop of the trial if the new drug or treatment is not effective in order to save time and resources. The probability of prestopping is high since most of the drugs currently being tested are high. So in our proposed two-stage screening design for phase II trials, percent of tumor size change endpoint is used as an initial

screening to select potentially effective agents within a short time interval followed by a second screening stage where PFS is estimated to confirm the efficacy of agents. This design can improve trial efficiency and reduce cost by early stopping the evaluation of an ineffective agent based on low percent of tumor size change. The second survival endpoint screening will substantially increase the success rate of follow up phase III trial by using the similar outcomes.

5.5.4 Critical Value in First Stage

We compared the two-stage phase II trials with different measurements of tumor shrinkages (continuous measurement of tumor size change vs. grouped tumor response rate). Moreover, we also use the simulation result to discuss the choice of primary test criterion in the first screening stage. As shown in section 4.4.3, the second stage success rate changes according to the selection of stage I significant levels decreasing according to significant level in first stage. This result may lead us to reflect the common idea for selection of candidate treatments. It is generally expected that a phase II trial with a relax criteria could increase the possibility of new drug discovery and avoid the omission from the phase II trial rejection, this assumption is not supported by our simulation result. A strict criterion in stage I screening in our phase II trial is considered to be more significant in practice, and tends to lead more satisfactory results in the further study.

6 CONCLUSION & FURTHER RESEARCH

It is the practical meaning that attracts our attention to start a research study in clinical trial study, with special interest in cancer clinical trial designs. As we know, cancer is one of the critical health issues in the world, even if the United States — a world leader in healthcare. After forty years of unremitting researches, the effective curative treatments for cancer patients are still in urgent need.

In developing new anti-cancer drugs, statistics is an indispensable and crucial element since clinical trial is a process that applies statistical inference theory into pharmaceutical research. This motivated us to focus on improving statistical designs of phase I and II trials to find solutions to several major obstacles in cancer clinical studies.

In this dissertation, to improve phase I trial efficiency, we proposed a hybrid design that considers three main competing interests: (1) preserve the safety of patients take part in the trial; (2) take into account of time factor allowing incomplete observations; and (3) fully utilize all toxicity information. We thus proposed a hybrid design for cancer phase I clinical trials — dose escalation method with overdose control using a normalized equivalent toxicity score system and time-to-event approach. The aim of design is to decrease the dose-finding trial duration, without impairing the characteristics of the EWOC design, especially the overdose control ability, as well as fully utilize all toxicity information to estimate MTD more accurately. Comparing with EWOC, EWOC-NETS, and EWOC-TITE, EWOC-NETS has the highest accuracy, followed closely by our proposed hybrid design — EWOC-NETS-TITE. If expressed in an inequality, it can be considered that the accuracy in MTD estimation is: EWOC-NETS > EWOC-NETS-TITE > EWOC. The “incompleteness” in observation is partially made up by the time-to-event approach and partially replenished by using more toxicity information. However, in the overall

evaluation, EWOC-NETS-TITE has a unique advantage that the trial duration is significantly reduced. There is one issue attracts our attention that in over-toxicity scenario EWOC-NETS derived designs tend to overestimate the MTD. One of the possible reasons could be the choice of toxicity profile. Therefore, expert inputs from preliminary experience and understanding of the testing drug is critical for phase I trials. And this issue would need further investigation in the later research.

For phase II trials, we proposed a novel two-stage design with double screening those improves the conventional phase II trial designs from the following aspects: (1) use continuous tumor size changes as endpoint in first screening stage; (2) utilize PFS as endpoint in second screening stage enhance subsequent phase III trial success rate; (3) double screening allows early termination which reduces the trial duration when no promising results present in stage I. The main advantage of our phase II two-stage design has shown in the result: Success rates in follow up phase III trials estimated by our proposed two-stage design vary from 24.14% to as high as 83.33%. Majority of the success rates calculated are above 50%, which performs better than all general methods for cancer phase II trials (50 – 60% of all phase III trials under investigation fail). Additionally, our proposed two-stage design has shown a series of very promising specificity and accuracy using simulation study. Simulation result shows that the proportions of true results that were successful detected by our proposed design were as high as 96.40%, and majority of the accuracies calculated were around 90%. However, the sensitivity which represents the probability of the design identifies candidate drugs which are effective is relative low comparing to Simon's two-stage design. Most sensitivity was around 40-50%, but some of them can be as low as 20%. While Simon's two-stage tends to have the opposite results. Our further research would focus on how to balance the relationship on "picking out" the candidate

treatment as wells as “avoiding” selection of inefficient agents and preventing waste of resource in follow up phase III trials. Moreover, Bayesian methodology implementation in phase II clinical trial designs is also our further research plan as it plays an increasingly prominent role in clinical trials. There are several topics that attract our attention for further research, such as sample size improvement, selection of appropriate prior informative distribution for Bayesian inferences, and choice of endpoints. Now the idea for next step study is to incorporate the Bayesian method to our proposed two-stage design.

Another well discussed topic in cancer clinical trials is the emergence of personalized anti-cancer drugs. Although progressions have been made to both diagnosis and treatment for cancers, there is still a long way to go. Today, forty years since the Nixon Administration declared the war on cancer; modern anti-cancer medications have saved millions of lives. However, a fact should be admitted that any one drug or treatment may not work for certain patients, even if it works for others. Or it may cause severe side effects on some people while not on the rest people. With better understanding of cancer biology today, it has been figured out that both genetics and environmental factors influence patients’ responses to anti-cancer treatments. It is the new trend in cancer drug development to tailor treatments to individuals. In this dissertation, we also proposed a phase I trial design that incorporates patient’s binary or continuous covariate effect in finding the dose-toxicity relationship. The advantage of our proposed model is it suggests a range of dose for according to patient’s characteristics. It could also accurately estimate the MTD for patients in different groups. Meanwhile the adjustment according to patients’ heterogeneity well controls the probability of overdosing a patient. In the further research direction, we would like to propose more complex model considering multiple patients’ characteristics with binary and continuous covariate incorporated.

REFERENCES

ACS Cancer Facts and Figures, 2012 [Report]. - Atlanta : American Cancer Society, 2012.

An MW Mandrekar SJ, Branda ME, Hillman SL, Adjei AA, Pitot HC, Goldberg RM, Sargent DJ. Comparison of continuous versus categorical tumor measurement-based metrics to predict overall survival in cancer treatment trials. [Journal] // Clin. Cancer Res. - 2011. - p. 17: 6592.

Arienti Chiara [et al.] Peritoneal carcinomatosis from ovarian cancer: chemosensitivity test and tissue markers as predictors of response to chemotherapy [Journal] // Journal of Translational Medicine. - 2011. - Vol. 9. - p. 94.

Babb James S and Rogatko Andre Patient specific dosing in a cancer phase I clinical trial [Journal] // Statistics in Medicine. - 2001. - Vol. 20. - pp. 2079-2090.

Babb James, Rogatko Andre and Zacks Shelemyahu Cancer Phase I Clinical Trials: Efficient Dose Escalation with Overdose Control [Journal] // Statistics in Medicine. - 1998. - Vol. 17. - pp. 1103-1120.

Bekele B N and Thall P Dose-finding based on multiple toxicities in a soft tissue sarcoma trial [Journal] // Journal of American Statistics Association. - 2004. - Vol. 99. - pp. 26-35.

Braun T M Generalizing the TITE-CRM to adapt for early- and late-onset toxicities [Journal] // Statistics in Medicine. - 2006. - Vol. 25. - pp. 2071-2083.

Brown S R [et al.] Designing phase II trials in cancer: a systematic review and guidance [Journal] // British Journal of Cancer. - 2011. - Vol. 105. - pp. 194-199.

Brown SR [et al.] Designing phase II trials in cancer: a systematic review and guidance [Journal] // British Journal of Cancer. - 2011. - Vol. 105. - pp. 194-199.

Burzykowski T, Molenberghs G and Buyse M The validation of surrogate endpoints by using data from randomized clinical trials: a case-study in advanced colorectal cancer [Journal] // Journal of Royal Statistical Society Series - Statistics in Society. - 2004. - Vol. 167. - pp. 103-124.

Buyse M [et al.] Progression-free survival is a surrogate for surival in advanced colorectal cancer [Journal] // Journal of Clinical Oncology. - 2007. - Vol. 25. - pp. 5218-5224.

Buyse M [et al.] Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer [Journal] // Lancet. - 2000. - Vol. 356. - pp. 373-378.

Buyse M [et al.] The validation of surrogate endpoints in meta-analyses of randomized experiments [Journal] // Biostatistics. - 2000. - Vol. 1. - pp. 49-67.

Chan Isaac S and Ginsburg Geoffrey S Personalized medicine: progress and promise [Journal] // Annual Review of Genomics and Human Genetics. - 2011. - Vol. 12. - pp. 217-244.

Chang Mark Adaptive dsign for clinical trials [Conference]. - France : International Symposium on Applied Stochastic Models and Data Analysis, 2005.

Chen T T Optimal three-stage designs for phase II cancer clinical trials [Journal] // Statistics in Medicine. - 1997. - Vol. 16. - pp. 2701-2711.

Chen Zhengjia [et al.] A novel toxicity scoring system treating toxicity response as a quasi-continuous variable in phase I clinical trials [Journal] // Contemporary Clinical Trials. - 2010. - Vol. 31. - pp. 473-482.

Chen Zhengjia [et al.] Methodology and Application of Adaptive and Sequential Approaches in Contemporary Clinical Trials [Journal] // Journal of Probability and Statistics. - 2012. - 2012 : Vol. 2012. - p. 20.

Chen Zhengjia, Tighiouart Mourad and Kowalski Jeanne Dose escalation with overdose control using a quasi-continuous toxicity score in cancer phase I clinical trials [Journal] // Contemporary Clinical Trials. - 2012. - Vol. 33. - pp. 949-958.

Cheung Y K and Chappell R Sequntial designs for phase I clinical trials with late-onset toxicities [Journal] // Biometrics. - 2000. - Vol. 56. - pp. 1177-1182.

Cheung Y K Coherence principles in dose-finding studies [Journal] // Biometrika. - 2005. - Vol. 92. - pp. 863-873.

Chow Shein-Chung and Chang Mark Adaptive design methods in clinical trials - a review [Journal] // Orphanet Journal of Rare Diseases. - 2008. - Vol. 3. - p. 11.

Cook Thomas and DeMets David L Review of Draft FDA Adaptive Design Guidance [Journal] // Journal of Biopharmaceutical Statistics. - 2010. - 6 : Vol. 20. - pp. 1132-1142.

Dhani Neesha [et al.] Alternate Endpoints for Screening Phase II Studies [Journal] // Clinical Cancer Research. - 2009. - 6 : Vol. 15. - pp. 1873-1882.

DiMasi J A and Grabowski H G Economics of new oncology drug development [Journal] // Journal of Clinical Oncology. - 2007. - Vol. 25. - pp. 209-216.

Durham S D, Fluournoy N and Rosenberger W F A random walk rule for phase I clinical trials [Journal] // Biometrics. - 1997. - Vol. 53. - pp. 745-760.

Eisenhauer E A [et al.] New response evaluation criteria in solid tumors: revised RECIST guidline (version 1.1) [Journal] // European Journal of Cancer. - 2009. - Vol. 45. - pp. 228-247.

Eisenhauer E A, O'Dwyer P J and Humphrey J S Phase I clinical trial design in cancer drug development [Journal] // Journal of Clinical Oncology. - 2000. - Vol. 18. - pp. 684-692.

Fanouriakis Antonis, Boumpas Dimitrios T and Bertsias George K Balancing efficacy and toxicity of novel therapies in systemic Lupus erythematosus [Journal] // Expert Review of Clinical Pharmacology. - 2011. - Vol. 4. - pp. 437-451.

FDA Draft Guidance Adaptive Design Clinical Trials for Drugs and Biologics [Online]. - 2010. - http://www.fda.gov/downloads/Drugs/.../Guidances/ucm201790.pdf.

Fleck Roman and Bach Daniel Trends in personalized therapies in oncology: the (venture) capitalist's perspective [Journal] // Journal of Personalized Medicine. - 2012. - Vol. 2. - pp. 15-34.

Friberg S and Mattson S On the growth rates of human malignant tumors: implications for medical decision making [Journal] // Journal of Surgical Oncology. - 1997. - Vol. 65. - pp. 284-297.

Gallo Paul [et al.] Adaptive designs in clinical drug development-an Executive Summary of the PhRMA Working Group. [Journal] // Journal of Biopharmaceutical Statistics. - 2006. - 3 : Vol. 16. - pp. 285-291.

Gehan E A The determination of the number of patients required in a preliminary and a follow-up of a new chemotherapeautic agent [Journal] // Journal of Chronic Disease. - 1960. - Vol. 13. - pp. 346-353.

Gill S [et al.] Progression-free survival as a primary endpoint in clinical trials of metastatic colorectal cancer [Journal] // Current Oncology. - 2011. - pp. 5-10.

Gray Robert [et al.] Phase II clinical trial design: methods in translational research from the genitourinary committee at the eastern cooperative oncology group [Journal] // Clinical Cancer Research. - 2006. - Vol. 12. - pp. 1966-1969.

Green S J and Dahlberg S Planned versus attained designs in phase II clinical trials [Journal] // Statistics in Medicine. - 1992. - Vol. 11. - pp. 853-862.

Gupta Umesh Chandra [et al.] Phase 0 clinical trials in oncology new drug development [Journal] // Perspective in Clinical Research. - 2011. - 1 : Vol. 2. - pp. 13-22.

Halabi Susan [et al.] Progression-free survival as a predictor of overall survival in men with castrate-resistant prostate cancer [Journal] // Journal of Clinical Oncology. - 2009. - Vol. 27. - pp. 2766-2771.

Ho Jeremy [et al.] Barriers in phase I cancer clinical trials referrals and enrollment: five-year experience at the Princess Margaret Hospital [Journal] // BMC Cancer. - 2006. - Vol. 6.

Hotte S J [et al.] Progression-free survival as a clinical trial endpoint in advanced renal cell carcinoma [Journal] // Current Oncology. - 2011. - pp. 11-19.

Ishizuka Naoki and Ohashi Yasuo The continual reassessment method and its applications: a Bayeisan methodology for phase I cancer clinical trials [Journal] // Statistics in Medicine. - 2001. - Vol. 20. - pp. 2661-2681.

Ivanova Anastasia and Flournoy Nancy Comparison of isotonic designs for dose-finding [Journal] // Statistics in Biopharmaceutical Research. - 2009. - Vol. 1. - p. 101.

Jemal Ahmedin [et al.] Cancer Statistics, 2008 [Journal] // CA: A Cancer Journal for Clinicians. - 2008. - 2 : Vol. 58. - pp. 71-96.

Joffe Steven and Miller Franklin G Rethinking risk-benefit assessment for phase I cancer trials [Journal] // Journal of Clinical Oncology. - 2006. - Vol. 24. - pp. 2987-2990.

Julka P K [et al.] Response assessment in solid tumours: a comparison of WHO, SWOG and RECIST guidelines [Journal] // The British Journal of Radiology. - 2008. - Vol. 81. - pp. 444-449.

Kang Seung-Ho and Ahn Chul The expected toxicity rate at the maximum tolerated dose in the standard phase I cancer clinical trial design [Journal] // Drug Information Journal. - 2001. - Vol. 35. - pp. 1189-1199.

Karrison T G [et al.] Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer [Journal] // Journal of the National Cancer Institute. - 2007. - Vol. 99. - pp. 1455-1461.

Korn E L [et al.] Clinical trial designs for cytostatic agents: are new approaches needed? [Journal] // Journal of Clinical Oncology. - 2001. - Vol. 19. - pp. 265-272.

Korn Edward L [et al.] A comparison of two phase I trial designs [Journal] // Statistics in Medicine. - 1994. - Vol. 13. - pp. 1799-1806.

Kramar A, Potvin D and Hill C Multistage designs for phase II clinical trials: statistical issues in cancer research [Journal] // British Journal of Cancer. - 1995. - Vol. 74. - pp. 1317-1320.

Kummar Shivaani [et al.] Phase 0 Clinical Trials: Conceptions and Msconceptions [Journal] // The Cancer Journal. - 2008. - 3 : Vol. 14. - pp. 133-137.

Lagakos S W Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable [Journal] // Statistics in Medicine. - 1988. - Vol. 7. - pp. 257-274.

Lavin P T An alternative model for the evaluation of antitumor activity [Journal] // Cancer Clinical Trials. - 1981. - Vol. 4. - pp. 562-457.

Lee J Jack and Feng Lei Randomized phase II designs in cancer clinical trials: current statust and furhter directions [Journal] // Journal of Clinical Oncology. - 2005. - Vol. 23. - pp. 4450-4457.

Leung Denis Heng-Yan and Wang You-Gan Isotonic designs for phase I trials [Journal] // Controlled Clinical Trials. - 2001. - Vol. 22. - pp. 126-138.

Lin Yong and Shih Weichung J Adaptive two-stage designs for single-arm phase IIA cancer clinical trials [Journal] // Biometrics. - 2004. - 2 : Vol. 60. - pp. 482-490.

Lin Yong and Shih Weichung Statistical properties of the treatment alorrighm-based designs for Phase I cancer clinical trials [Journal] // Biostatistcs. - 2001. - 2 : Vol. 2. - pp. 203-215.

Louvet C [et al.] Correlation between progression free survival and response rate in patients with metastatic colorectal carcinoma [Journal] // cANCER. - 2001. - Vol. 91. - pp. 2033-2038.

Murgo Anthony J [et al.] Designing Phase 0 Cancer Clinical Trials [Journal] // Clinical Cancer Research. - 2008. - Vol. 14. - pp. 3675-3682.

Naggara O [et al.] Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptred aneurysms [Journal] // American Journal of Neuroradiology. - 2011. - Vol. 32. - pp. 437-440.

NCI Cancer Clinical Trials [Online]. - 2013. - http://www.cancer.gov/cancertopics/factsheet/Information/clinical-trials.

NCI Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03 [Online]. - June 14, 2010. - http://www.eortc.be/services/doc/ctc/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf.

Newell D R Pharmacologically based phase I trials in cancer chemotherapy [Journal] // Hematology. - 1994. - Vol. 8. - pp. 257-275.

NIH Cost of Cancer Care [Online]. - National Institute of Health, June 2012. - http://progressreport.cancer.gov/doc_detail.asp?pid=1&did=2011&chid=105&coid=1026&mid=.

Offit Kenneth Personalized medicine: new genomics, old lessons [Journal] // Huamn Genetics. - 2011. - Vol. 130. - pp. 3-14.

O'Quigley John, Pepe Margaret and Fisher Lloyd Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer [Journal] // Biometrics. - 1990. - 1 : Vol. 46. - pp. 33-48.

Parulekar Wendy R and Eisenhauer Elizabeth A Phase I trial design for solid tumor stuides of targeted, non-cytotoxic agents: theory and practice [Journal] // Journal of the National Cancer Institute. - 2004. - Vol. 96. - pp. 990-997.

Pivot X [et al.] Response rates: a valuable signal of promising activity? [Journal] // Cancer Journal. - 2009. - Vol. 15. - pp. 361-365.

Potter D Phase I studies of chemotherapeutic agents in cancer patients: a review of the designs [Journal] // Journal of Biopharmaceutical Statistics. - 2006. - Vol. 16. - pp. 58=79-604.

Ratain M J [et al.] Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents [Journal] // Journal of the National Cancer Institute. - 1993. - Vol. 85. - pp. 1637-1643.

Roberts T G [et al.] Trends in the risks and benefits to patients with cancer participating in phase I clinical trials [Journal] // Journal of American Medicine Association. - 2004. - Vol. 292. - pp. 2130-2140.

Robinson William T Innovative Early Development Regulatory Approaches: expIND, expCTA, Microdosing [Journal] // Clinical Pharmacology & Therapeutics. - 2008. - Vol. 83. - pp. 358-360.

Rogatko A [et al.] Patient-specific dose adjustment in the cancer clinical trial setting [Journal] // Pharmaceutical Medicine. - 2008. - Vol. 22. - pp. 345-350.

Rosenberger W F and Haines L M Competing designs for phase I clinical trials: a review. [Journal] // Statistics in Medicine. - 2002. - Vol. 21. - pp. 2757-2770.

Royston P, Altman D G and Sauerbrei W Dichotomizing continuous predictors in multiple regression: a bad idea. [Journal] // Statistics in Medicine. - 2006. - Vol. 25. - pp. 127-141.

Rubinstein Lawrence V [et al.] Design issues of randomized phase II trials and a proposal for phase II screening trials [Journal] // Journal of Clinical Oncology. - 2005. - Vol. 23. - pp. 7199-7206.

Saad E D [et al.] Progression-free survival as surrogate and as true endpoint: insights from the breast and colorectal cancer literature [Journal] // Annals of Oncology. - 2010. - Vol. 1. - pp. 7-12.

Sachs R K, Hlatky L R and Hahnfeldt P Simple ODE Models of Tumor Growth and Anti-Angiogenic or Radiation Treatment [Journal] // Mathematical and Computer Modelling. - 2001. - Vol. 33. - pp. 1297-1305.

Scher H I and Heller G Picking the winners in a sea of plenty [Journal] // Clinical Cancer Research. - 2002. - Vol. 8. - pp. 400-404.

Seymour L [et al.] The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the National Cancer Institute investigational drug steering committee [Journal] // Clinical Cancer Research. - 2010. - Vol. 16. - pp. 1764-1769.

Shuster J Optimal two-stage designs for single arm phase II cancer trials [Journal] // Journal of Biopharmaceutical Statistics. - 2002. - Vol. 12. - pp. 39-51.

Siegel Rebecca, Naishadham Deepa and Jemal Ahmedin Cancer Statistics, 2012 [Journal] // CA: A Cancer Journal for Clinicians. - 2012. - 1 : Vol. 62. - pp. 10-29.

Simon R [et al.] Accelerated titration designs for phase I clinical trials in oncology [Journal] // Journal of the National Cancer Institute. - 1997. - Vol. 89. - pp. 1138-1147.

Simon R M [et al.] Clinical trial designs for the early clinical development of therapeutic cancer vaccines [Journal] // Journal of Clinical Oncology. - 2001. - Vol. 19. - pp. 1848-1854.

Simon R Optimal two-stage designs for phase II clinical trials. [Journal] // Controlled Clinical Trials. - 1989. - Vol. 10. - pp. 1-10.

Simon R, Wittes R E and Ellenberg S S Randomized phase II clinical trials [Journal] // Cancer Treatment Report. - 1985. - Vol. 69. - pp. 1375-1381.

Skipper H and Schabel F Jr. Quantitative and cytokinetic studies in experimental tumor systems [Book]. - Philadelphia : Lea and Febiger, 1982. - 2nd.

Storer Barry E Design and analysis of phase I clinical trials [Journal] // Biometrics. - 1989. - Vol. 45. - pp. 925-937.

Stylianou Mario and Flournoy Nancy Dose finding using the biased coin up-and-down design and isotonic regression [Journal] // Biometrics. - 2002. - Vol. 58. - pp. 171-177.

Tan S B and Machin D Bayesian two-stage designs for phase II clinical trials [Journal] // Statistics in Medicine. - 2002. - Vol. 21. - pp. 1991-2012.

Thompson E I Application of restricted sequential design in a clinical protocol [Journal] // Cancer Treatment Reports. - 1980. - Vol. 64. - pp. 399-403.

Tighiouart M and Rogatko A Dose finding in drug development [Book Section] // Dose Finding in Oncology-Parametric Methods. - New York : Springer, 2006.

Tighiouart Mourad, Cook-Wiens Galen and Rogatko Andre Incorporating a patient dichotomous characteristic in cancer phase I clinical trials using escalation with overdose control [Journal] // Journal of Probability and Statistics. - 2012. - Vol. 2012. - p. 10.

Todd Susan Sequential designs for monitoring two endpoints in a clinical trial [Journal] // Drug Information Journal. - 1999. - Vol. 33. - pp. 417-426.

Tourneau Christophe Le, Lee J. Jack and Siu Lillian L Dose Escalation Methods in Phase I Cancer Clinical Trials [Journal] // Journal of the National Cancer Institute. - 2009. - 10 : Vol. 101. - pp. 708-720.

Walker I and Newell H Do molecularly targeted agents in oncology have reduced attrition rates [Journal] // Nature Reviews Drug Discovery. - 2009. - Vol. 8. - pp. 15-16.

Wang Y [et al.] Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development [Journal] // Clinical Pharmacology and Therapeutics. - 2009. - Vol. 86. - pp. 167-174.

Wason J MS and Mander A P The choice of test in phase II cancer trials assessing continuous tumor shrinkage when complete responses are expected [Journal] // Statistical Methods in Medical Research. - 2011. - pp. 1-11.

Whitehead John The design and analysis of sequential clinical trials [Book]. - UK : John Wiley & Sons Ltd., 1997. - 2nd.

WHO Cancer Fact Sheet [Online]. - 2013. - http://www.who.int/mediacentre/factsheets/fs297/en/.

WHO WHO handbook for reporting results of cancer treatment [Book]. - Geneva, Switzerland : World Health Organization Offset Publication, 1979.

Yang S and Zhao Y Testing treatment effect via empirical likelihood using weighted log rank tests [Journal] // Statistics and Probability Letters. - 2007. - Vol. 77. - pp. 1385-1393.

Yothers G Toward progression-free survival as a primary endpoint in advanced colorectal cancer [Journal] // Journal of Clinical Oncology. - 2007. - Vol. 25. - pp. 5253-5154.

Yuan Z, Chappell R and Bailey H The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach [Journal] // Biometrics. - 2007. - Vol. 63. - pp. 173-179.

Yuan Zhilong and Chappell Rick Isotonic designs for phase I cancer clinical trials with multiple risk groups [Journal] // Clinical Trials. - 2004. - Vol. 1. - pp. 499-508.

Zandvliet Anthe S [et al.] Two-stage model-based clinical trial design to optimize phase I development of novel anticancer agents [Journal] // Investigational New Drugs. - 2010. - Vol. 28. - pp. 61-75.

Zhao Y and Zeng D Recent development on statistical methods for personalized medicine discovery [Journal] // Fronters of Medicine. - 2013. - Vol. 7. - pp. 102-110.