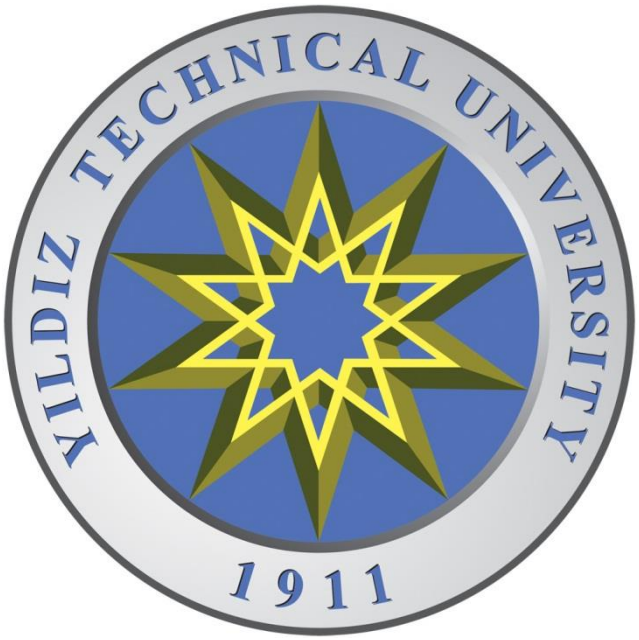


**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

**THESIS**

**NOVEMBER, 2017**

**REPUBLIC OF TURKEY**
**YILDIZ TECHNICAL UNIVERSITY**
**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# INVESTIGATION OF PHASE I DOSE FINDING METHODS: BAYESIAN METHODOLOGY FOR PHASE I STUDIES

**EFEHAN ULAŞ**

**PhD THESIS**
**DEPARTMENT OF STATISTICS**
**PROGRAM OF STATISTICS**

**ADVISER**
**ASSOC. PROF. DR. FİLİZ KARAMAN**

**ISTANBUL, 2017**

**REPUBLIC OF TURKEY**

**YILDIZ TECHNICAL UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# INVESTIGATION OF PHASE I DOSE FINDING METHODS: BAYESIAN METHODOLOGY FOR PHASE I STUDIES

The thesis submitted by Efehan ULAŞ in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 03.11.2017 in Department of Statistics, Statistics Program.

**Thesis Adviser**

Assoc. Prof. Dr. Filiz KARAMAN

Yıldız Technical University

**Approved By the Examining Committee**

Assoc. Prof. Dr. Filiz KARAMAN

Yıldız Technical University _____

Prof. Dr. Ali Hakan BÜYÜKLÜ

Yıldız Technical University _____

Yrd. Doç. Dr. Elif Çiğdem ALTUNOK

Yeditepe University _____

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Yıldız Technical University _____

Prof. Dr. Ahmet Mete ÇİLİNGİRTÜRK

Marmara University _____

## ACKNOWLEDGEMENTS

It is my greatest respect to express my thanks to those who helped me during my Ph.D. I would like to thank my adviser Dr. Filiz KARAMAN for teaching, encouragement, and support. I appreciate the valuable guidance and assistance given by her. I am very thankful to all valuable members of the jury who provided constructive inspiration and criticism in my dissertation.

I must acknowledge my parents for their unconditional love and confidence in me.

November 2017

Efehan ULAŞ

## TABLE OF CONTENTS

## LIST OF SYMBOLS

| | |
|---|---|
| $\beta_e$ | Elimination rate parameter |
| $B_{10}$ | Bayes factor for models |
| C(t) | Drug concentration at any time t |
| $C_{E50}$ | Concentration of the drug at Emax |
| $E_{max}$ | Maximum effect of the drug |
| d | Dose of a drug |
| α | Model parameter |
| λe | Boundary of escalation |
| $x_i$ | Range of the dose levels |
| λd | Boundary of de-escalation |
| $\Omega_j$ | Prior information of the subjects |
| $Y_j$ | Response of the subject |
| V | Volume of the distribution |
| $\psi(x_i, \alpha)$ | Dose response model |
| $\tilde{x}_j$ | Prior mean of the toxicity probabilities |
| $\phi_T$ | Target toxicity probability |
| $\tilde{\psi}_j$ | Posterior mean of the toxicity probability |
| $\bar{\psi}_j$ | Toxicity probability at each dose |
| $\varepsilon_1$ | Small fraction |

## LIST OF ABBREVIATIONS

| | |
|---|---|
| BCRM | Bayesian Continual Reassessment Method |
| mTPI-2 | Bayesian Interval Dose-Finding Design Addressing Ockham's Razor |
| BIC | Bayesian Information Criteria |
| BMA-CRM | Bayesian Model Averaging with Continual Reassessment Method |
| BOIN | Bayesian Optimal Interval Designs |
| CDK4/6 | Caclin Dependent Kinase Retinoblastoma |
| CRM | Continual Reassessment Method |
| DLT | Dose Limiting Toxicity |
| FDA | Food and Drug Administration |
| MTD | Maximum Tolerated Dose |
| mTPI | Modified Toxicity Probability Interval Method |
| RP2D | Recommended Phase II Dose |
| H9N2 | Seroprevalence of Avian Influenza |
| SHR6390 | Small Molecular Oral Patent |
| TDL | Toxicity Dose Limitation |
| TPI | Toxicity Probability Interval |
| UPM | Unit Probability Mass |
| VEGF | Vascular Endothelial Growth Factor |

## LIST OF FIGURES

Page

## LIST OF TABLES

Page

# ABSTRACT

## INVESTIGATION OF PHASE I DOSE FINDING METHODS: BAYESIAN METHODOLOGY FOR PHASE I STUDIES

Efehan ULAŞ

Department of Statistics

PhD. Thesis

Adviser: Assoc. Prof. Dr. Filiz KARAMAN

Clinical trial is a scientific study designed to examine whether new potential treatments are safe and effective. Despite its name, a clinical trial is not conducted in a laboratory, in fact, the clinical trial experience is very similar to a regular doctor visits. Healthy volunteers or patients with the illnesses that are being studied participate in the trial. The aim is to gather enough evidence to understand if a medicine works and it is safe. Participating in a clinical trial means being part of the advancement of health care and science. During the clinical trial, the participants often have the opportunity to access promising new treatments that may be more effective than the current standard of care. Safety of the potential treatment is the first determined treatment during phase I trials, but it is continuously monitored through all phases. This thesis reviews the most popular and used phase I dose response methods, and explores key limitations of these methods, and introduces a comparative simulation study that has different model structures and prior distributions in the Continual Reassessment Method (CRM).

The common phase I methods are the 3+3 design, A+B design, Continual Reassessment Method (CRM), Bayesian Model Averaging Continual Reassessment Method (BMA-CRM), Bayesian Optimal Interval Designs (BOIN), Modified Toxicity Probability Interval Method (mTPI ) and a Bayesian Interval Dose-Finding Design Addressing Ockham's Razor (mTPI -2). These methods are used in the clinical trials to select a true maximum tolerated dose (MTD). In the first part of the thesis, these methods were compared to two different stories and twelve different scenarios. After examining and comparison of each scenario, which methods were more effective and efficient in selecting the correct MTD was concluded. According to the results, CRM, BMA-CRM, mTPI and mTPI-2 were the best performing methods in our simulation runs. In the second part of the thesis, a different story and eight scenarios are implemented. The prominent methods in the first part were compared with the CRM method where the model structure and prior distribution were different.

Overall, in designs where the model structure is hyperbolic tangent and prior distribution is uniform, the CRM calculated the selection probability of the correct MTD higher than the others. On the other hand, in designs where the model structure is logit and prior distribution is lognormal, the CRM calculated the selection probability of the correct MTD lower than other CRM designs. In addition, the BMA-CRM produced very effective results if the difference between the correct MTD dose and the previous-subsequent dose is greater. Moreover, the mTPI and mTPI-2 designs can produce better results in the case of where the target toxicity of the trial is not included in the study. In conclusion, more reliable and applicable results for phase I dose finding trials are produced by the BMA-CRM and CRM, when the model structure and prior distributions are different, in our study. As a result, the model-based designs performed much better than the rule-based designs.

**Key words:** Dose response models, clinical trials, phase I, maximum tolerated dose, dose finding, 3+3 design

# ÖZET

## FAZ I DOZ BULMA YÖNTEMLERİNİN İNCELENMESİ: FAZ I ÇALIŞMALARI İÇİN BAYESCİ YÖNTEM

Efehan ULAŞ

İstatistik Anabilim Dalı

Doktora Tezi

Tez Danışmanı: Doç. Dr. Filiz KARAMAN

Klinik deneme, yeni potansiyel tedavilerin güvenli ve etkili olup olmadığını incelemek üzere tasarlanmış bir bilimsel araştırmadır. Adının klinik deneme olmasına rağmen, bu çalışmalar herhangi bir laboratuarda yapılmaz. Genellikle bir doktor ziyaretine benzemektedir. Sağlıklı gönüllüler veya araştırılan hastalığı bulunan bireyler denemeye katılırlar. Bu denemelerdeki amaç ilacın hastalığı tedavi edip etmediğini anlamak için yeterli delil veya veriyi toplamaktır. Bir klinik araştırmaya katılmak, sağlık ve bilim alanındaki ilerlemenin bir parçası olmak demektir. Klinik araştırmalarda, katılımcılar sıklıkla mevcut standartlara kıyasla daha etkili umut verici yeni tedavilere erişme fırsatına sahip olurlar. Potansiyel tedavinin güvenliği ilk olarak faz I çalışmaları sırasında tespit edilir, ancak her aşamada sürekli izlenir. Bu tez, en popüler ve en çok kullanılan faz I doz yanıt yöntemlerini inceleyerek, bu yöntemlerin eksik yanları araştırmakta ve sürekli yeniden değerlendirme (CRM) methodunda farklı model yapıları ve önsel dağılımları kullanıldığında doğru dozu bulmadaki seçim olasılıklarının nasıl değiştiğini karşılaştırmalı bir simülasyon çalışması ile göstermektedir.

Genel olarak bilinen faz I doz bulma yöntemleri; 3 + 3 tasarımı, A + B tasarımı, Sürekli Yeniden Değerlendirme Metodu (CRM), Bayesci Model Ortalama Sürekli Yeniden Değerlendirme Metodu (BMA-CRM), Bayesci Optimum Aralık Tasarımı (BOIN), Modifiye Edilmiş Zehirlilik Olasılık Aralığı Yöntemi (mTPI) ve Ockham'ın Tıraşlığını ele alan Bayesci Aralıklı Doz Bulma Tasarımıdır (mTPI -2). Bu yöntemler, doğru maksimum tolere edilebilir dozu (MTD) seçmek için klinik araştırmalarda kullanılır. Tezin ilk bölümünde, bu yöntemler iki farklı gerçek öykü ve on iki farklı senaryo ile karşılaştırılmıştır. Bu karşılaştırmalar sonucunda her senaryo incelenmiş ve hangi yöntemlerin doğru MTD seçiminde daha etkili olduğu bulunmuştur. Sonuçlara göre,

simülasyon çalışmalarımızda CRM, BMA-CRM, mTPI ve mTPI-2 en iyi performansı gösteren yöntemler olmuşlardır.Tezin ikinci bölümünde farklı bir gerçek öykü ve sekiz senaryo kullanılmıştır. Birinci bölümdeki önemli yöntemler, model yapısı ve önsel dağılımın farklı olduğu CRM yöntemi ile karşılaştırılmıştır.

Genel olarak, model yapısının hiperbolik tanjant ve önsel dağılımın tekdüze olduğu tasarımlardaki CRM, doğru MTD'nin seçim olasılığını daha yüksek hesaplamıştır. Öte yandan, doğru MTD'nin seçim olasılığını, model yapısının logit ve önsel dağılımın lognormal olduğu tasarımlardaki CRM, diğer CRM tasarımlarından daha düşük hesaplamıştır. Buna ek olarak, eğer doğru MTD dozu ile bir önceki ve sonraki doz arasındaki fark büyük ise BMA-CRM çok iyi sonuçlar vermiştir ve araştırmanın hedef toksisitesinin çalışmaya dahil edilmediği durumlarda mTPI ve mTPI-2 tasarımlarını kullanmak daha iyi sonuçlar verebilir. Sonuç olarak çalışmamızda, faz I doz bulma deneyleri için daha güvenilir ve uygulanabilir sonuçlar, BMA-CRM ve CRM'nin model yapısı ve önsel dağılımının farklı olduğu dizaynlar tarafından üretilmiştir. Ayrıca model tabanlı tasarımlar, algoritma tabanlı tasarımlardan çok daha iyi performans göstermiştir.

**Anahtar Kelimeler:** Doz cevap modelleri, klinik denemeler, faz I, maksimum tolere edilebilir doz, doz bulma, 3+3 dizayn

# CHAPTER 1

# INTRODUCTION

## 1.1 Literature Review

Clinical trials are studies that play a critical role on how new medical approaches work on human bodies and help researchers to identify appropriate dosage of the drug. Moreover, they aim to detect, treat or manage known and unknown diseases or medical conditions. According to Friedman et al. [1], it is a prospective study for determining the effect of treatment in human. The clinical trials are one of the stages of long and diligent processes. Experts have been working for many years to understand the effects of the new treatments as well as their side effects. Due to these side effect risks, the clinical trials are started with a small group of people to reduce any possible damages. Although there is no certainty that the clinical trials will result in favorable treatment, the participating patients provide significant contributions to the future treatments. To achieve the desired results, medical doctors, researchers, and patients should work together with a care and loyalty for finalizing the clinical trials.

In general, the clinical trials are classified as four consecutive phases [2]. The trial design for each phase is a complex process, and usually requires a close collaboration among academic institutions, medical centers or hospitals, pharmaceutical companies, public organizations and regulatory agencies. Figure 1.1 illustrates the clinical trial process of the four phases. Those four phases are explained in the following sections.

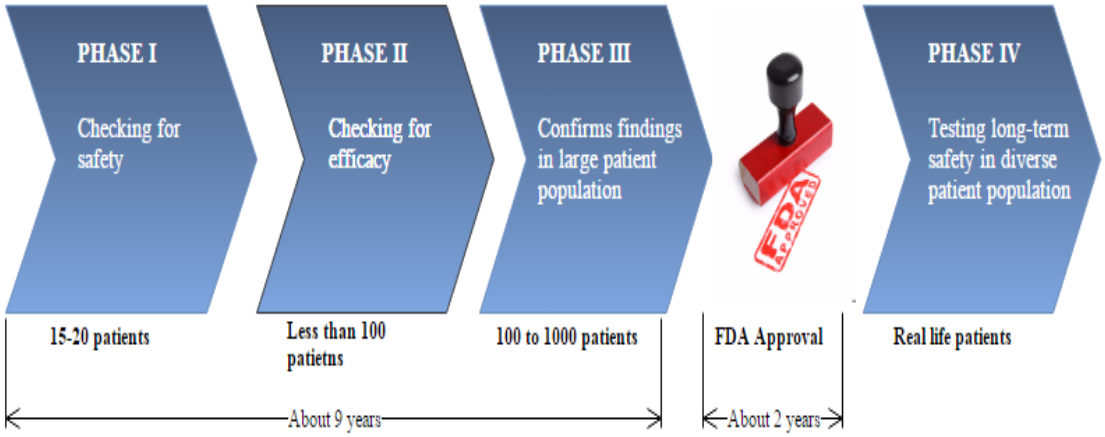


Figure 1.1 Clinical Trial Process

In common, the clinical trials are separated into two groups. The first group is called trial with a control group and the second trial is called a group-free trial. In cases where the clinical trials do not include a comparison between the applied treatment and the other treatment, or if the patient's enrollment date and the treatment control dates do not match, those experiments are called uncontrolled trials. The uncontrolled trials are the experiments that an investigator indicates his experience with used drug and treatment. In contrast, the controlled clinical trials may include standard treatments or a placebo for direct comparison so that the difference between the clinical outcomes of the experimental treatment can be objectively assessed [3]. The clinical trials are called to have internal validity if the difference between the observed treatment group is real (no biased or systematic error). Generally, randomized, double blind (clinician and patients do not know the identity of treatment), placebo-controlled trials affect high levels of internal validity. In the clinical trials, external validity determines whether the outcome of the research is widespread. It wouldn't be relevant if the internal validity was uncertain. The external validity can be increased with patients' eligibility criteria.

The clinical trials are the most effective approach for comparing and examining experimental drugs, medical treatments, or clinical interventions on humans. Furthermore, the outcome of the clinical trials has an important impact on clinical practice. Well designed and appropriately conducted clinical trials are very powerful tools for the discovery of the new drugs and the development of the existing drugs.

The first controlled clinical trial is the study that was conducted by Lind (1753) for the scurvy disease. In this study, twelve patients were divided into six groups as two persons in each group [4]. In similar conditions, the same diet was applied to all

patients. Two subjects were given a quart of cider a day; two patients received vitriol potions three times a day; two took two tablespoons of vinegar, three times a day; two patients with the poorest condition were put under a course of seawater; two of them had two oranges and a lemon per day; and the remaining two patients received meatballs three times a day. The most effective result was perceived from the orange and lemon treatment. It is clear that this study lacks some basic characteristics of the modern clinical trials. For example, the patients were not properly randomized; two patients with the poorest condition were treated with sea water and the study was not blinded or masked; that is, both the patient and the investigator knew what the treatment was. Therefore, there might have been bias in the selection and the other confounding effects in this study.

The clinical trials are designed with constant sample width and equal randomization. In the equal randomization, the patients are assigned with equal probabilities to each treatment method. On the other hand, adaptive randomization is more ethical approach than the equal randomization because it tends to assign more patients to more intensive treatments based on the available evidence [5]. Generally, adaptive designs are studies that evaluate the response of a small number of patients in early phase studies. The adaptive design enhances the efficiency of use of the patient data by combining the data from learning phase and the data from confirmatory phase.

The adaptive design can provide a stronger outcome from the patient’s data in the study and shorten the duration of drug development [6]. The adaptive design can be used as a tool in the planning of the clinical trials in difficult experimental situations. In all cases, the intermediate analysis and the proposed design types must be defined in the study protocol. The use of Adaptive Design shows that the statistical method controls a predefined Type 1 error [7], that the correct prediction and confidence intervals for the treatment effect are present, and that methods for evaluating the homogeneity of results from different phases are preplanned.

Today's clinical trials use much more modern methods and statistical information than before. Now, the design of the clinical trials has become more advanced with the management of the adaptive statistical methods. For instance, the first trial was conducted by the University of Texas, MD, known as BATTLE (Biomarkers-Integrated Approaches of Targeted Therapy for Lung Cancer Elimination) [8]. This study consists of four parallel phase 2 studies for patients with advanced stage non-small cell lung

cancer. The trial is continued to learn the treatment effects on the biomarker profiles of the patients.

Secondly, highly anticipated multi-agent experiments (I-SPY 2) were conducted [9]. This adaptive neoadjuvant phase 2 study was performed on woman with newly diagnosed locally advanced breast cancer. The aim was to investigate whether the targets of the investigated drug combinations were better than the targets of the standard chemotherapy, which is merely chemotherapy.

### 1.1.1 Drug Development

The focus of the drug research and development process is the patients. The purpose of drug development is to help the patients overcome their illnesses, and improve their quality of life [10]. The drug development process is designed to ensure that innovative medicines are effective, safe and have the treatments that can be brought to the patient's use as soon as possible. The first stage of the drug development is the discovery of the best targets that can prevent a disease [11]. These targets are usually the proteins in the microorganism that cause the disease-related proteins in the patient's body. The challenge here is to identify the role of the disease process in determining which treatments are more relevant and more important [12]. At this stage, cellular networks and protein networks need to be characterized.

A single protein can send messages to many proteins. This can sometimes be due to multiple metabolic pathways that can affect function. Understanding how these metabolic pathways work and interact helps to select appropriate targets for a drug. Focusing on metabolic targets helps to better understand the mechanism of the disease [13]. When this information and the desire to find a solution to unmet medical drugs combined, the discovery of biological targets can be realized. In drug discovery, high-throughput screening attempts to find chemical compounds or biological substances that can be linked to targets by methods such as computer-aided design [14].

If a compound can change the target to affect the disease, it is called a hit. These are developed again with efficacy and safety, including drug candidates. The discovery of a medicine and the introduction of medicines require an investment of approximately two billion dollars [15], with approximately fifteen years of research and clinical development. Only one of the ten thousand hit tests tested in the first stages of the drug discovery could become a medicine offered to the service of the medicine. At the end of

the preclinical period, some additional investigations are carried out to determine that drug candidates are safe for the patients and that they use the appropriate pharmacokinetic properties, such as absorption metabolism, appropriate for humans [16]. These experiments are carried out with extraordinary sensitivity in order to reduce the risk of human suffering.

Animals play an important role in the drug discovery process. Although many research and development studies can be done experimentally and through computers, complex disease mechanisms can often be understood by the animal studies. In addition, governments and regulatory agencies want drugs tested on animals before humans [17]. The clinical study programs consist of several phases that assess drug safety, efficacy, and effectiveness, and these phases are explained in the next sections.

### 1.1.2 Phase I

Phase I trials are small trials that recruit only 15-20 patients. In the phase I trials, pharmaceutical operators are applied to humans for the first time [18]. These trials are conducted in the clinic to observe the side effects of a candidate drug on volunteers [19]. If the new treatment is effective on such disease, the answers to the following questions are investigated:

- What is the safest dose level for patients?
- Which side effects are occurred during treatment?
- How do the patients cope with the drug?
- Can treatment treat the current diseases?

The objective of these trials is to determine any drug tolerance and interaction description properties of pharmacokinetics and to identify intervals of the dose levels and side effects. The researchers who lead the trials also try to determine the dose level of the drug and find out the best way to give the drug (oral, injection etc.). The primary goal of the phase I trial is to understand the drug tolerance in volunteers and identify dose limiting toxicities (DLTs). To do so, the maximum tolerated dose (MTD) is needed to be determined. The MTD is the highest dose of a drug which does not cause any unacceptable side effects. So, the MTD is the dose where the probability of dose limiting toxicity (DLT) is equal to predefined level. The MTD can be represented as the

dose at where the probability of the DLT is the probability that is targeted $P(\text{DLT})= \theta$. Here $\theta$ is the probability that targeted.

In the phase I studies, it is expected that the toxicity will be increased when the dose level is high. In such cases, the MTD is supposed to be the most promised dose [20]. However, this hypothesis might be inaccurate when the doses are lower than the MTD.

### 1.1.3 Phase II

The phase II trials follow the completion of the phase I trials. The effectiveness of a drug is called drug's efficacy in the phase II trials. These trials evaluate potential efficacy and focus on the therapeutic effects of the drug. In the phase II trials, the limited numbers of patients are treated with a specific dose of the drug, and it is important to assess how well the drugs continue safety assessments.

The effect of the new drug plays a critical role in the phase II studies because the process goes through the validator tests to make long-term studies if the new drug has a positive effect on the treatment [21]. The phase II trials may involve single-arm studies or there may be multi-arm studies between randomly selected patients and different treatment modalities. In comparison to the phase I studies, the phase II trials are relatively large. It usually varies from 30 to 100 samples. Phase II experiments are sometimes classified as phase IIa or phase IIb studies. Phase IIa trials are conducted for future studies to ensure that they are not active against the disease, and are generally used for comparison purposes when there is no standard improvement. Phase IIb trials are similar to phase III trials and are used to study the effects of the drug. These tests should be compared with at least one standard treatment. In general, phase I and phase II trials are carried out separately in succession and each requires a significant planning and inspection phase [22].

### 1.1.4 Phase III

Drugs which are effective and appropriate for further study in phase II trials are tested in phase III trials. A Larger group of the patients are used in phase III clinical trials. In this period, improvement activity, side effects, and advantage/disadvantage ratio are determined and compared with the other drugs are made. By using the same protocol, this can be done in different hospitals simultaneously.

Phase III trials are the most costly comparative studies used to assess the efficacy of drugs because they involve large numbers of patients and require long follow-up times. These studies involve between 100 to 1000 patients and last for two to four years or longer due to recruitments [23]. Calculation of the sample size in a Phase III experiment is the most critical part of the experimental design [24], [25].

Within the framework of the hypothesis tests, first type error probability $\alpha$, second type error probability $\beta$ and the effect size should be determined. If the sample size is not large enough, the trial can be misleading to find an effective medication because the statistical test cannot reach the appropriate level of importance due to lack of power. On the other hand, if the sample size is overestimated, large resources and efforts will be made. More importantly, the drug development can be delayed because it makes it difficult to test patient records.

### 1.1.5 Phase IV

Phase IV trials are the studies that take a very long time to approval of the intervention [26]. In this phase, effectiveness and safety of an intervention are monitored. Subject's tolerance may change after some time. Because of that, it is important to continue monitoring the patients in a long period of time. Since the number of subjects is less for early trials, safety problems may occur in this phase. Therefore, the monitoring for a long time is needed. Also, it is important to examine the outcomes because the quality of life of the patients may be affected by the treatment. Generally, phase IV trials are used for post-marketing surveillance which means watching the drug's long-term effects. This help to monitor unseen side effects in the early phases. These studies are carried out once the drug license has been obtained then it can be sold in the market.

## 1.2 Aim of the Study

Recently, several doses finding methods are developed. However, the comparison of those designs and the traditional designs are not intensive. The first goal of this study is to compare the most used phase I dose finding methods and to determine which one is performing better results. The second goal of this dissertation is to show how different model structures and priors in the CRM design are effective in selecting the correct dose level as the MTD.

## 1.3 Hypothesis

Clinical trials are the most effective approach for comparing and examining experimental drugs, medical treatments, or clinical interventions on humans. Safety of the potential treatment is the first determined treatment during phase I trials, but it is continuously monitored through all phases. In this study, we introduced a comparative simulation study that has different model structures and prior distributions in the Continual Reassessment Method (CRM).

## 1.4 Structure of the Thesis

Chapter 1 thoroughly introduces the clinical trials and its concept which are used in this thesis. It gives detailed reviews of key components of clinical trials. For a better understanding, of the role of the statistical problems arising from the clinical trials, the most important fundamentals of the clinical trials are introduced in this chapter.

Chapter 2 reviews the literature on the most used designs in the phase I clinical trials. It begins by introducing the adaptive and the sequential methods. Accordingly, it investigates the methodology of the most used model and the rule based phase I designs. Therefore, Chapter 2 explains each design in detail. Moreover, the setup and limitation of the methods are compared with each other.

The stories of the simulation studies by using real life scenarios are presented in Chapter 3. The scenario I and Scenario II are used to compare the most used phase I methods. Afterwards, the scenario III is used to compare the best-performed methods in the scenario I and II with the CRM (when the model selection and prior distribution is different).

In Chapter 4, simulation studies using different scenarios are presented. In the first and the second simulation studies, the best-performed methods are underlined. The findings of these simulation studies are compared with the CRM design when it has different model structures and priors. Finally, some major gains of using different model structures and priors in the CRM design is highlighted.

Chapter 5 summarizes the major contributions of this dissertation, the results of previous chapters and discusses possible future research.

# CHAPTER 2

# METHODS IN PHASE I CLINICAL TRIALS

Development of candidate medicines is a long, difficult and expensive process. In each phase of the trial, statistical methods play a critical role in the development of successful medicines. In the phase I, the trial is conducted in a small sample size because statistical approaches are indispensable and important for the clinical trial analysis [42]. In the last decade, group sequential and adaptive designs become more popular in the clinical trials. The development of these designs significantly changed the traditional drug discovery paradigm [43]. With these designs, the analyses can be performed on accumulating data. In case of any toxicity, the trial can be stopped early. Therefore, the statistical approaches help to conduct a trial with minimum cost, time and maximum treatment effect for the patients.

## 2.1 Terms in Phase I Clinical Trials

This section provides brief information about basic statistical concepts in order to understand the problems that may be encountered in the clinical trials, and then the detailed information is given about pharmacokinetic and pharmacodynamics modeling.

### 2.1.1 Protocol

Every clinical trial has a study protocol that describes the appropriate plan and the features of the research. The protocol is the document that describes the entire study in a comprehensive way. A protocol includes general information about the treatment conditions, disease, the purpose of the study, the procedures used to evaluate the safety and efficacy of the drugs, statistical designs and methods, objectives and eligibility criteria. The protocol also includes potential risks in the treatment, benefits of the treatment administered to human subjects, targeted patient population, the standard of

the dose level and the time of the drug administration [4]. The content of the protocol which is recommended by WHO (World Health Organization) contains the following sections:

- General, rationale and background information
- Objectives of the study
- Design of the study
- Methodology
- Safety Considerations
- Data management and statistical analysis
- Expected outcomes of the study
- Dissemination of results and publication policy
- Time period of the project
- Regulatory, administrative and legal obligations.

The protocol requires statistical evaluations to ensure the integrity of the trial design. These statistical evaluations are; sample size, probability model, efficacy and toxicity monitoring, stopping criterion for futility and superiority, statistical methods and programs for interim data, assessment of the possible deviations from the original study plan in the procedure. Any changes made in the study after the protocol must be added to the protocol.

### 2.1.2 Power and Sample Size

Determination of the sample size in the clinical trials is an important part of the clinical protocol. In a trial where the sample size is too large, the trial may reach its goal before the end of the study, and some experimental units might be unnecessarily included in the study. On the other hand, in a study where the sample size is too small, the chances of reaching the goal of the research will be very low. For this reason, in a clinical trial, the sample size should be large enough to give reliable answers to the questions being addressed [27]. To determine the appropriate sample size, the primary variable, the test statistic for the hypothesis to be tested, the null hypothesis, the alternative hypothesis, the probability of type I error and type II error should be specified. In addition, the primary dependent variable indicating the efficacy of the test must be clearly defined. Endpoint variables might be binary variables such as patients' response to treatment or continuous variable such as blood pressures, measurements of cholesterol levels.

Different types of endpoint variables require different statistical methods to calculate the sample size [28].

In the clinical trials, variance and clinical trial design parameters should be known for predicting the sample size. If these parameters are unknown, or the predictions are made from the literature and the pilot studies, it will lead to sampling error. Estimates are treated as mass values and this approach includes sampling error. Therefore, the size of the sample obtained might be misleading. In this case, the use of the Bayesian approach is recommended [29], because the Bayesian approach can be used for both the precision and the power analysis methods. The estimation of the sample size is approached as a decision problem and a loss or utility function is used. Bayesian estimates can be obtained by using appropriate loss functions instead of unknown parameters.

### 2.1.3 Blinding

Blinding is concealment of a group assignment by one or more people participating in the clinical research trials, mostly randomly controlled. Randomization reduces the differences between the treatment groups at the beginning of the experiment to the greatest extent and then has no effect to interfere with the different treatments of the experimental groups or the differential evaluation of the results. This may lead to biased estimates of the treatment effects [30]. Bias may occur in the clinical trials, as the patients generally expect to get the most recent and effective treatments while the doctors wish to participate in a successful trial. Both the patient and the doctor want the effects of a new treatment to be more positive, which can cause some side effects to be reported incompletely. For this reason, it is critical to neutralize such bias by masking the treatment identity so that the trial participants are blinded to the nature of the treatment [31].

At each stage of a trial, there are various people who can be biased. These include patients, principal investigators, physicians, surgeons, local researcher coordinators, principal laboratories reporting scans or blood samples, trial statisticians, and committees such as the decision-making or data monitoring and safety committee. For preventing bias, each of these groups can be blinded. The use of the blinding in those studies strengthens the results of that research. In the clinical trials, the blinding can be performed in four different ways:

- **Un-blinded:** All parties are aware of the treatment the participant receives.
- **Single blinded:** Only the participant is unaware of the treatment.
- **Double blinded:** Both the participant and the researcher are unaware of the treatment.
- **Triple blinded:** The patients, the researchers, and the data analysts are unaware of the treatment.

The blinding of the trial requires careful planning and continuous monitoring to ensure that the blinding is maintained. It is also ensures that the safety of the patient and the validity of the trial results are not exceeded. It is also vital that all study protocols are clearly documented that they show which patients are blinded and how they are blinded in the study because it can have a significant effect on the value of the study results [31]. Figure 1.2 shows the allocation of the type of blinding.

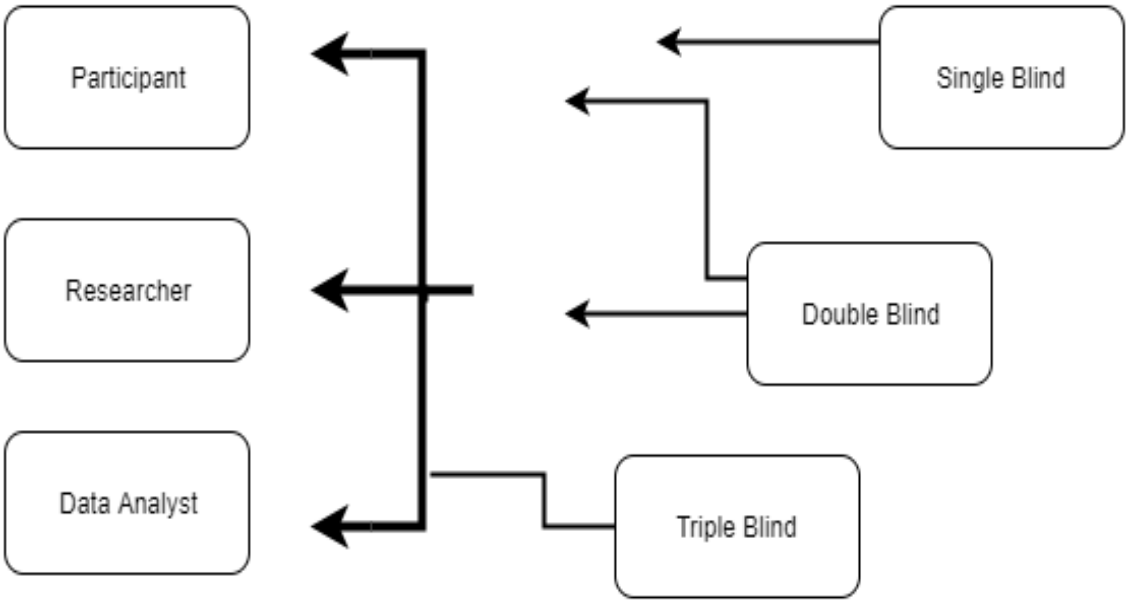


Figure 2.2 Type of Blinding

### 2.1.4 Randomization

Randomization is an estimated allocation of the patients to a specific treatment strategy in the clinical trial. When a large number of patients are included, the simple randomization will balance the groups in an experiment in terms of patient characteristics and the other factors that produce biased results. It can then be assumed that the remaining differences in efficacy or safety outcomes among the groups are due to the effects of different treatment strategies or random errors. The randomization is often used in experimental designs. It is important that the sequence to be traced in the

experiment is random. When a decision is made to control certain variables at certain levels; it is a fact that the effects of the other variables that cannot be controlled. The randomization of the order of the experiment is intended to neutralize these uncontrolled variables. The randomization also allows the researchers to think as if they were free from measurement errors, and so it is preferred in many statistical studies.

The creation of a randomization design can be accomplished by using one of the several procedures. Once a design and allocation method has been agreed, the rules must be followed throughout the study. The most used randomization methods are:

- Simple randomization
- Block randomization
- Stratified randomization
- Minimization randomization

There are also some other hybrid randomization methods. However, the ones mentioned above are the most used ones [32], [33].

### 2.1.5 Crossover Design

Parallel experiment design is a one-arm experiment design in which the participants are distributed to random treatment groups. In parallel experiment designs, variability in terms of response and endpoint is separated into two different forms; inter-individual and intra-individual variability. Intra-individual variability is the time-dependent change in the same participant. A crossover experiment design is an experimental design type in which two or more test strips are applied in a specific order to each of a certain number of participants [34]. This design is the most common design used in many clinical and pharmacological trials to compare different assays with each other. Researchers often prefer crossover designs due to budget constraints by adding new patients to the trial, difficulties in finding patients to participate in the trial, and time constraints in the specific training of each new patient in the trial.

In the two-arm crossover design, each patient is evaluated twice. The new method is applied to the first group (treatment A) and classical approach or placebo is applied to the control group (treatment B). To eliminate any residual effects, the washout period between two treatments should be applied. In the second phase of the study, treatment B is applied to the new individuals in the first group whereas the new method (treatment

A) is applied to the second group [35]. This two-arm crossover design can be seen in Figure 2.2.

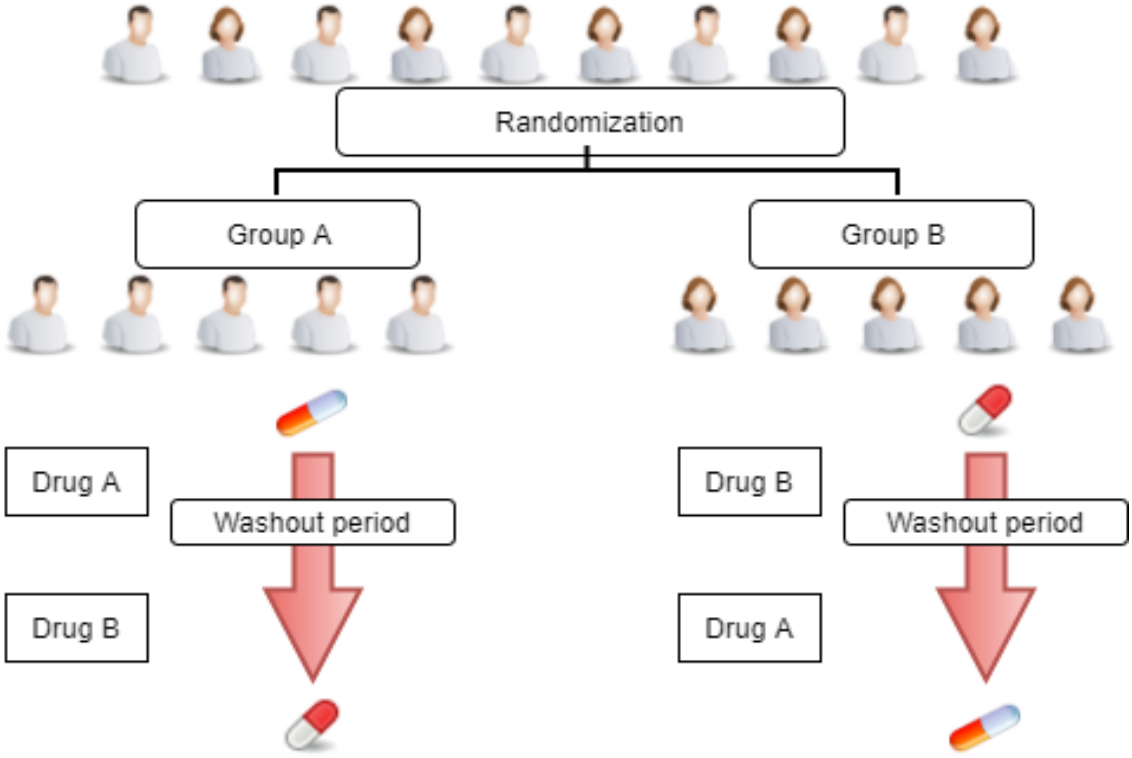


Figure 2.3 Two-arm Crossover Design

### 2.1.6 Data Collection and Adverse Events

Clinical trial data are obtained from medical examinations, laboratory test results, patient interviews, and questionnaires. Extreme care should be given for collecting data so that trial results can be interpreted in the best way. The values of dependent variables are obtained from the test results. The dependent variables are clinically relevant and should be responsive to primary and secondary questions. The dependent variables can be binary, continuous or time to event measurement [36]. In general, an answer to the primary problem must be defined by a single dependent variable. If more than one dependent variable is present in the experiment, it might be more difficult to generalize the findings.

Patients and researchers are generally concerned about adverse events or adverse side effects in drug-related studies. During the trial, toxicity should be continuously monitored. The frequency of administration of the drug and the level of toxicity should be minimized when dose quantities are selected [37]. If adverse events or side effects are encountered, the trial should be stopped.

### 2.1.7 Pharmacokinetics and Pharmacodynamics

Pharmacokinetics (PK) and pharmacodynamics (PD) properties of drugs should be well understood in preclinical studies and early phase trials. PK examines the processes of the absorption, distribution, transformation, and body-wasting of drugs in the body by establishing mathematical models [38]. It investigates how organism's interaction with the drug. On the other hand, the PD investigates the effects of drugs on living organisms and their mechanisms of action [39]. It takes care of what medicines do to the organism. Figure 2.3 shows the difference between the PK and PD clearly.

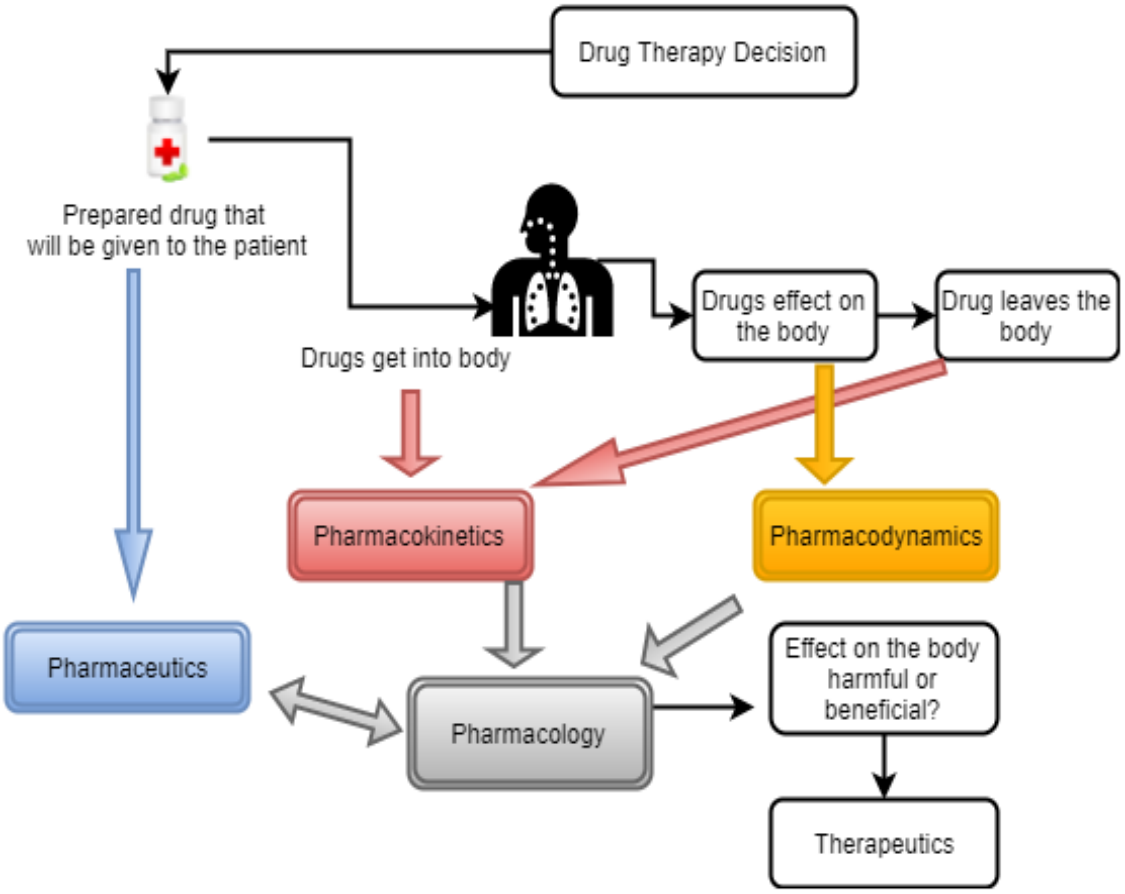


Figure 2.4 Drug Concentration Diagram

Pharmacokinetics model shows the relationship between drug blood concentration and time period. If the drug concentration at any time *t* is represented by C(t), the model will be as follows [40],

$$C(t) = \frac{d}{V}\exp(-\beta_e t) \tag{1.1}$$

Here, d is the dose of the drug, V is the volume of the distribution, and $\beta_e$ is the elimination rate parameter. Assuming that the drug is administered as a single oral dose, the model of the relationship between drug blood concentration and time period will be as follows [40],

$$C(t) = \frac{\beta_a d}{(\beta_e - \beta_a)V}\{\exp(-\beta_e t) - \exp(-\beta_a t)\} \tag{1.2}$$

The most commonly used model drug concentration-effect profiles is the sigmoid $E_{max}$ model. $E_{max}$ model has a monotonically increasing curve that is flattened when it reaches the maximum drug action level. This model also known as Hill equation [41]. So the form of the therapeutic effects of the drug can be seen below;

$$E = E_{max} = \frac{C^{\alpha}}{(C_{E50})^{\alpha} + C^{\alpha}} \tag{1.3}$$

Here, $E_{max}$ is the maximum effect of the drug and C is the concentration of the drug. So that $C_{E50}$ indicates the concentration of the drug when $E_{max}$ is 50%.

## 2.2 Maximum Tolerated Dose and Initial Dose

The phase I clinical trials are initiated as the first step in drug testing on human subjects if pre-clinical evidence of progression is seen to prevent disease. The main goal of the phase I trials is to determine the recommended dose level for phase II trials. The phase I trials are concerned about the choice of the initial dose level, the efficacy of the trial, the rate of the effect of the dose, the likelihood of target toxicity, the number of patients and the adequacy of the experimental design. The goals of the phase I study are to determine the maximum tolerated dose (MTD), calculate safety and tolerability, and calculate the pharmacodynamics and pharmacokinetics of the new drug. The MTD is the dose with the likelihood of the toxicity closest to the target toxicity ratio determined by the investigators. The recommended phase II dose level should be at a MTD or below a MTD [44]. The toxicity dose limitation (TDL) is called stopping the treatment when the level of drug toxicity is at a high level.

In the early stages of drug development, the knowledge of the proper dosage of the new drug and its human subjects were inadequate. For this reason, the most toxicity but tolerable dose for the MTD should be observed with subsequent doses applied during the study. For the phase I studies, the sample size is small, and typically between 15-20

subjects. In the phase I, dose finding is adaptable. The patients are included in the study in turn and are usually treated in groups. The trial is initiated from the dose set by the investigator or the lowest dose [45]. If the undesirable number of patients encounters the TDL at the administered dose level, the next group will be treated with a lower dose, if the dose is tolerable, the next group will be treated with a higher dose. The optimal dose level is assigned to each patient group based on the data obtained from the experiments by the end of the trial. A general dose finding situation is shown in Figure 2.4. From the six increasing doses, the fourth dose seems to have a target toxicity ratio of 25%.

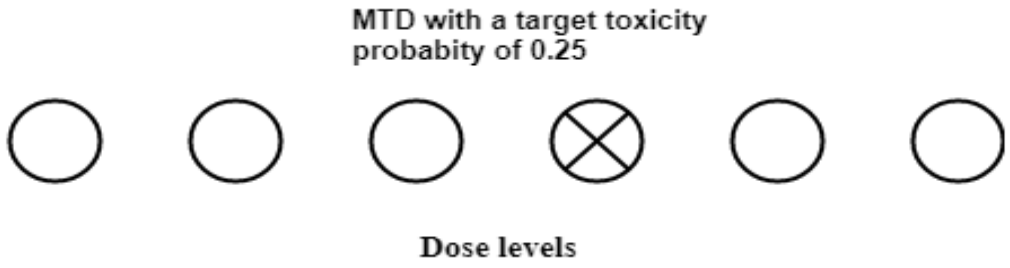


Figure 2.5 Simple Dose Levels for the MTD

To determine the MTD, the toxicity data obtained by dose increase and reduction is collected from the treated patients. It is expected that toxicity will appear immediately after the treatment. For safety reasons, the dose of the drug is gradually increased starting from the lowest level. Thus, the MTD is likely to cover the entire dose range.

## 2.3 Adaptive and Sequential Methods

In the clinical trials, the most important objective is to evaluate the efficacy of a test treatment against a control. To do so, a well-prepared study protocol is essential. The protocol should be valid to provide an unbiased assessment of the treatment safety. The statistical methods should be employed for assessment of the effect [28]. In the light of this, the adaptive and sequential methods have become popular in the last decade. In the group sequential method, the dataset is evaluated sequentially as they are assembled, and a trial is monitored sequentially for stopping the study when the result has enough evidence [42]. Moreover, it provides the number of stages, critical values, sample size and a stopping criterion to reject or accept the null hypothesis at each interim stages. In every interim stage, all data which is collected up to that point is analyzed and maximum likelihood test statistics and the standard errors are calculated. Then, the test

statistics are compared with the calculated critical values of the sequential design and the decision of stopping or continuing the trial can be made. If the trial continues to the last stage, the null hypothesis is either rejected or continued.

The group sequential methods provide flexibility in the design phase of the trial. However, the modification in conduct phase is not allowed. For instance, ad hoc increase in sample size cannot be made but adaptive designs overcome this limitation [46].

The adaptive designs present a method that allows modification during the trial. Uncertainty of the decisions that are made when conducting the trial can be addressed in the adaptive designs [47]. These features make the adaptive designs very attractive to clinical researchers and sponsors. In 2006, FDA released a Critical Path Opportunities List that presents many biomedical projects. This list consists of six research areas which are clinical trial designs, bioinformatics, public health, manufacturing, biomarkers, and pediatrics. Two of those research areas supported for improving the drug development (clinical trial design, biomarkers) and these projects call for advancing innovative trial methods such as adaptive designs to improve clinical development [40]. The other published documents on the use of the adaptive designs can be found on the FDA website.

In the thesis, we considered the only adaptive methods due to their flexibility during the trial. Therefore, this chapter will review the theory of adaptive design methods.

## 2.4 3+3 Design

The most popular choice of the phase I dose finding method among clinicians in the clinical trials is 3+3 design [48]. The reason is that the idea of the 3+3 design is simple and it is easy to implement. The 3+3 design is a rule-based design and it proceeds with cohorts of three patients. DLT (dose limiting toxicity) is the most important factor in the assessment of the doses. In this method, the first cohort is treated at a starting dose level (lowest dose level) and other subjects are enrolled in cohorts with increasing dose levels. The research goes further to the higher doses which depend on the assessment of the previous doses. The algorithm of the 3+3 design is described as follows,

a) Assess the toxicity of 3 treated patients at jth dose level

b) If any of the 3 patients' dose does not reach DLT, the dose level is increased to j+1 and then back to step 1

c) If one of three patients encounter DLT, 3 patients are treated with the same dose level j and the process is as follows:

- If one of six patients encounters DLT, the j dose level is increased to the dose level j+1 if j is non-exceeding MTD
- If two out of six patients meet the DLT, the experiment is terminated and a sub-six dose level is identified as j-1 MTD.
- If more than 2 patients meet with DLT, the current dose level is considered as exceeded the MTD and 3 patients will be treated at the j-1 level and less than 6 patients will be treated at this dose level.

d) If two or three patients meet DLT (dose level exceeded MTD), 3 patients will be treated at a j-1 dose level, based on the fact that less than 6 patients are treated at a j-1 dose level.

In the dose-finding studies, dose escalation should proceed cautiously to avoid exceeding the MTD. Thus, the patients are protected from high toxicity dose levels [49]. On the other hand, doses should be increased rapidly to avoid treating a large number of patients with ineffective doses below the MTD. Generally, the MTD is found to be the highest dose with the possibility of toxicity lower than 33% [50]. Figure 2.5 illustrates the process of the 3+3 design briefly.

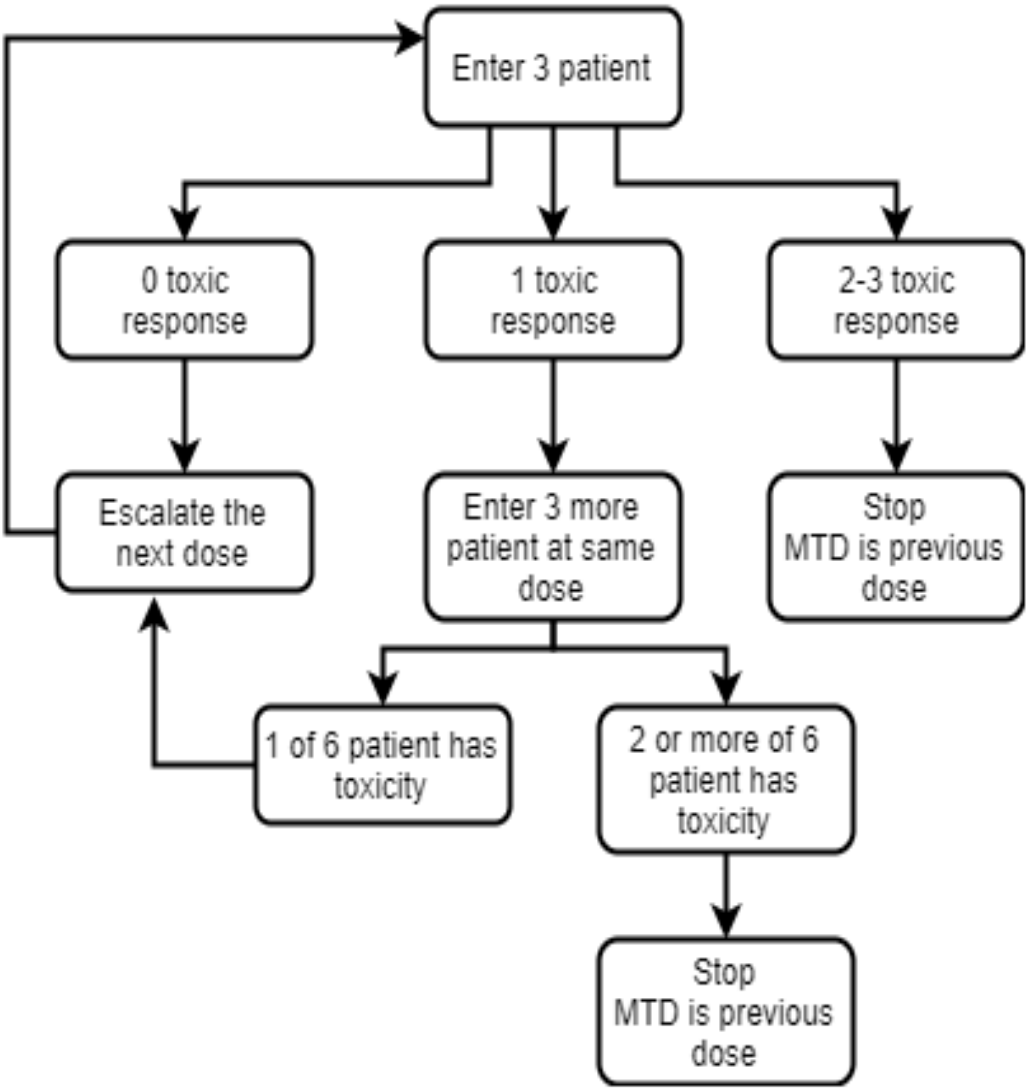


Figure 2.6 Diagram of 3+3 Design

## 2.5 A+B Design

The name of the A+B design comes from the two integers A and B which represents the number of patients at a given dose level. Five parameters are required for the full specification of the design; A, B, C, D, and E, respectively[51] [52]. The 3 + 3 design can be expanded to the A + B design. The cohort of the patients in the group may not always be three [53]. In this situation, the A+B design can be appropriate for further analysis. The parameters A - E are defined as follows: A is the number of the patients assigned to a dose in a first cohort, B is the number of the patients assigned a dose in a second cohort, C is the minimum number of the DLTs needed out of A patients to assign B, D is the maximum number of the DLTs required to make more assignments from the patients A to B; otherwise it stops the trial or de-escalate, E specifies the

maximum number of the DLTs allowed in A + B patients [54].. The traditional 3+3 design can be obtained from the A+B design by using (A, B, C, D, E) as (3, 3, 1, 1, 1) [51], [53]. Figure 2.6 shows the diagram of the A+B design without de-escalation.

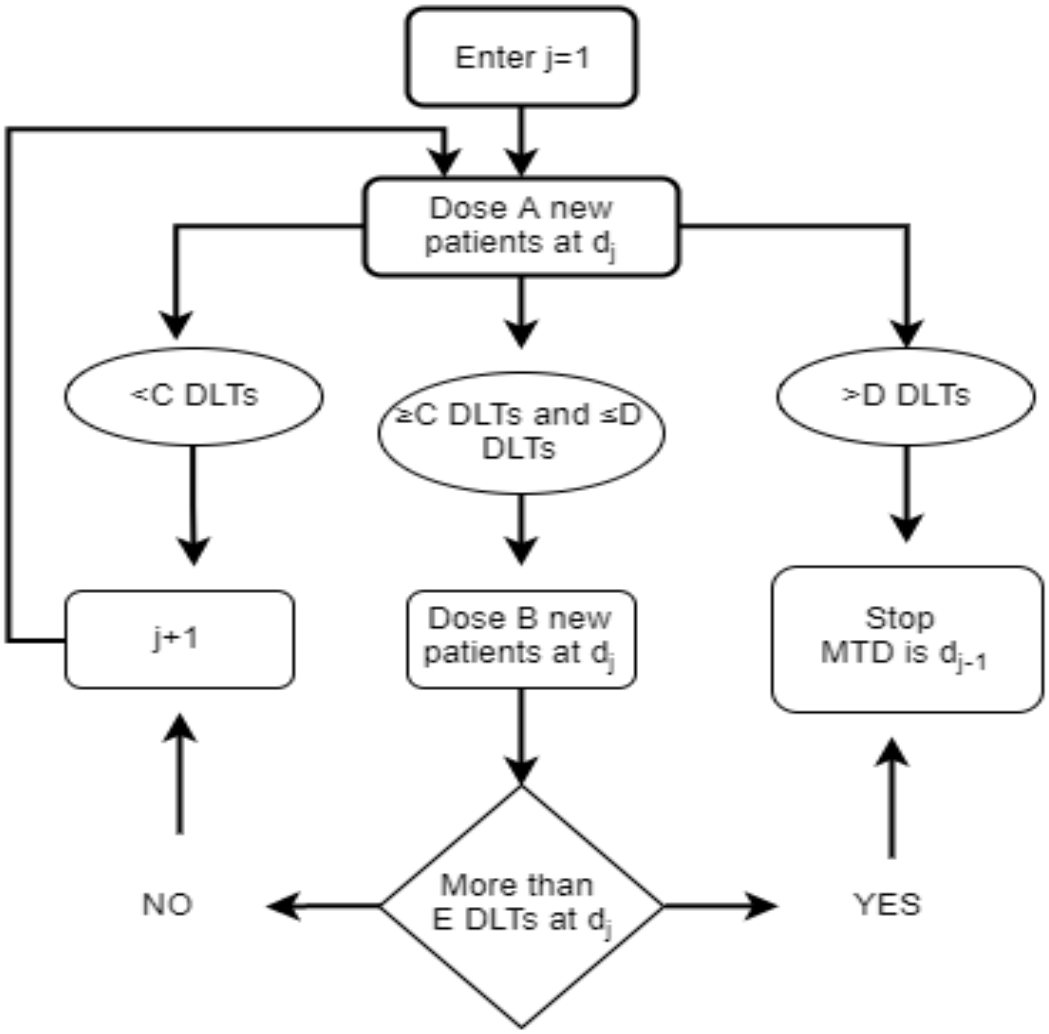


Figure 2.7 Diagram of the A+B design

## 2.6 Continual Reassessment Method

The Continual Reassessment Method (CRM) is one of the model-based approaches of the dose-finding methods in the drug development and it was the first proposed by Quigley et al. [55]. In algorithm-based dose finding methods, modeling information from the other doses is not used. Thus, it only considers the data observed from current dose level. In contrast, model-based dose finding methods assume a precise parametric model for the dose-toxicity curve. In particular, the continual reassessment method (CRM) links likelihood of the toxicity at each dose level with a pre-determined toxicity probability via a one-parameter model. While the toxicity data are collected, the CRM

continually updates the estimation of the toxicity probabilities of all doses. Each new patient cohort is transferred to the most appropriate dose based on the updated toxicity probabilities and the MTD occurs when the entire sample size is used. There are several extended versions of the traditional CRM. For example, Goodman et al. [56] suggests that the patients can be included in trials with cohorts of two and three patients, rather than enrolling the patients individually in the CRM. Garrett-Mayer [57] recommends applying a two-parameter logistic model in complex dose-response interactions. Yin and Yuan [58] introduces the BMA-CRM design that uses multiple initial guesses in the trial. Cheung and Chappell [59] presented a new version of the CRM that allows enrolling the patients in the trial before the previous patients complete the trial. Faries [60] modified the traditional CRM that allocates the initial trial patients at the lowest dose and the next patients at the highest dose that does not exceed the MTD. Korn et al. [61] recommends stopping the trial after six patients have already been enrolled at the next recommended dose. Gasparini and Eisele [62] introduces a CRM design that uses the toxicity probabilities directly. Piantadosi et al. [63] suggests cooperating with physicians who are familiar with the trial in order to understand dose-response characteristics. In this dissertation, the traditional CRM is considered.

In general, toxicity is assumed to be monotonically increasing depending on the dose. A detailed content of the CRM framework is explained below.

Suppose that a range of dose levels, $x_i (i = 1, \dots, k)$. The response of the *j*th subject where ( *j*= 1,...,k):

$$Y_j = \begin{cases} 1, & if\ toxicity \\ 0, & if\ no\ toxicity \end{cases}$$

Hyperbolic-tangent dose-response model proposed by [55]: $\psi(x_i,\alpha)$,

$$E(Y_j) = \tilde{\theta} = \psi(x_i, \alpha) = [\frac{\tanh(x_i) + 1}{2}]^{\alpha} \quad (2.1)$$

Power dose-response model: $\psi(x_i,\alpha)$,

$$\psi(x_i, \alpha) = {x_i}^{\alpha} \quad (2.2)$$

Logistic model with fixed intercept, c, where c is often taken to be 3: $\psi(x_i,\alpha)$,

$$\psi(x_i, \alpha) = \frac{\exp(3 + \alpha x_i)}{1 + \exp(3 + \alpha x_i)} \tag{2.3}$$

Prior distribution for the parameter α for j*th* subject: $f(\alpha, \Omega_j)$ where $\Omega = \{y_1, \dots, y_{j-1}\}$ and, $\int_0^\infty f(\alpha, \Omega_j) da = 1$, where j= 1,...n. The probability of toxicity at dose level $x_i$ using information from previous response: (i=1,...,k),

$$\theta_{ij} = \int_0^\infty \psi(x_i, \alpha) f(\alpha, \Omega_j) da \tag{2.4}$$

The posterior estimates of α are very intensive. Alternatively, approximate mean response probability:

$$\theta`_{ij} = \psi\{x_i, \mu(j)\} \text{ and } \mu(j) = \int_0^\infty \alpha f(\alpha, \Omega_j) da \tag{2.5}$$

Above equation is an estimate of the probability of a DLT for each dose level. Some important hints for CRM are:

- In dose-response model, $\psi(x_i, \alpha)$, $x_i (i = 1, \dots, k)$ are pre-determined dose levels and α is a model parameter.
- The unit exponential distribution g(α)=exp(-α) is the most used prior distribution in literature.
- The recommended percent for the DLT is usually between 20-33 [64].

### 2.6.1 Bayesian Update

Let assume the selected dose as x(j) and to find $f(\alpha, \Omega_{j+1})$ with j*th* patient response:

$$\phi(x(j), y_j, \alpha) = [\psi(x(j), \alpha)]^{y_j} [1 - \psi(x(j), \alpha)]^{1-y_j} \tag{2.6}$$

A Bayesian update is implemented by using a prior distribution, g(α), on the model parameter. The posterior distribution of α is calculated with the formulation below.

$$f(\alpha, \Omega_{j+1}) = \frac{f(\alpha, \Omega_j) \phi(x(j), y_j, \alpha)}{\int_0^\infty f(u, \Omega_j) \phi(x(j), y_j, u) du} \tag{2.7}$$

So suppose $y_j$ subjects experienced the DLT among the $n_j$ subjects treated at dose level j. If we take the power dose-response model, the likelihood function will be as below:

$$f(\alpha, \Omega_{j+1}) = \frac{g(a)\prod_{l=1}^{j}\phi(x(l), y_l, \alpha)}{\int_0^{\infty} g(u)\prod_{l=1}^{j}\phi(x(l), y_l, u)\,du} \tag{2.8}$$

Let $\phi_T$ be the target toxicity probability and $f(\alpha, \Omega_j)$ denote a prior distribution for α, for instance α~lognormal (μ,$\sigma^2$) where model parameter is power model, the posterior mean of the toxicity probabilities ($\tilde{\psi}_j$) can be calculated from the (2.8),

$$\tilde{\psi}_j = \int {x_j}^{\exp(\alpha)} \frac{\prod_{l=1}^{j}({x_j}^{\exp(\alpha)})^{y_j}\,(1-{x_j}^{\exp(\alpha)})^{1-y_j} f(\alpha, \Omega_j)}{\int_0^{\infty}\prod_{l=1}^{j}({x_j}^{\exp(\alpha)})^{y_j}\,(1-{x_j}^{\exp(\alpha)})^{1-y_j} f(u, \Omega_j)du} \tag{2.9}$$

The prior for α can be taken as Gamma, Uniform, Lognormal. However, Gamma(1,1) is the prior that used mostly in the literature [65]. In addition, the dose response models that can be used in the CRM are power, hyperbolic tangent and logistic dose response models. However, power dose-response model is the most used model in studies. Dose-response models that have used in this thesis can be seen in Table 2.1.

Table 2.1 Dose-Response Models

| Dose-Response Model | $\psi(x_i, \alpha)$ |
|---|---|
| **Power** | ${x_i}^{\alpha}$ |
| **Hyperbolic Tangent** | $[\frac{\tanh(x_i)+1}{2}]^{\alpha}$ |
| **Logistic** | $\frac{\exp(3+\alpha x_i)}{1+\exp(3+\alpha x_i)}$ |

### 2.6.2 Skeleton of the CRM

The skeleton is a very important set of probability in the CRM that directly affects the MTD. Our aim is to choose a set of $\{x_j\}$, which can reflect the true dose toxicity. The selection of the skeleton depends on the clinician's experience [66]. So it can be subjective. Use of the different skeleton in the CRM can cause different operating characteristics[58].

Prior mean toxicity probabilities can be defined as ( $\tilde{x}_1, \dots, \tilde{x}_j$ ) instead of $(x_1, \dots, x_j)$. This is also known as the skeleton of the CRM. So we can calculate the skeleton $(x_1, \dots, x_j)$ from,

$$\tilde{x}_j = E\left(x_j^{\exp(\alpha)}\right) = \int x_j^{\exp(\alpha)} f(\alpha) da, \text{ where j=1,…,J} \quad (2.10)$$

It is important to note that $(x_1, \ldots, x_j)$ is the main component of the CRM. This initial guess probabilities are not the same with the prior distributions as in the Bayes theorem.

### 2.6.3 Dose Finding Algorithm

In the CRM, the subjects are usually treated with a cohort of 3. The process of the CRM is as follows.

a. The lowest dose or determined dose level by the clinician is given to the first cohort.
b. $j^{cur}$ represents the current dose level. The posterior means of the toxicity probabilities are obtained from the observed data and all other doses are under consideration. The posterior means of the toxicity probabilities shown as $\tilde{\psi}_1, .., \tilde{\psi}_j$ and let $\phi_T$ be the target toxicity probability. To find the dose level, $j^*$, that has the toxicity probability closest to $\phi_T$,

$$j^* = \underset{j \in (1, \ldots, j)}{argmin} \left|\tilde{\psi}_j - \phi_T\right| \quad (2.11)$$

- If $j^{cur} > j^*$, decrease the dose $j^{cur} - 1$
- If $j^{cur} < j^*$, increase the dose $j^{cur} + 1$
- Else, dose level stays at the same level

c. When the maximum sample size is reached, the dose with the closest possible toxicity to $\phi_T$ is selected as maximum tolerated dose.

## 2.7 Bayesian Model Averaging Continual Reassessment Method

BMA-CRM is the model-based dose finding method that aims to identify the MTD. This method pre-specifies multiple sets of mean toxicity probabilities [58]. The popular model-based dose escalation method the CRM requires predetermination of the toxicity probability at each dose. This can be discretionary, and it can lead to different design properties. To handle this problem, Yin and Yuan [58] propose the use of multiple parallel CRM models each with a different set of predetermined toxicity probabilities. For each CRM model, they placed discrete probability masses. Those probability

masses are used as the prior model probability and they obtained posterior probabilities by the Bayesian model averaging approach [67]. Moreover, Madigan and Raftery [68] pointed that by averaging the all considered models, predictions can be better than single model based approaches. In the light of this, the dose increase and dose decrease are determined based on the target toxicity rate and the dose toxicity probabilities are estimated by the BMA.

To understand the theory behind the BMA-CRM, let $(M_1, .., M_K)$ be the models corresponding to each skeleton, $\{(x_{11}, \dots, x_{1J}), \dots, (x_{K1}, \dots, x_{KJ})\}$, so we can define the model $M_k$, (k=1,...,K), which uses the k*th* initial guesses,

$$\psi_{kj}(\alpha_k) = {x_{kj}}^{\exp(\alpha_k)}, \quad j = 1, \dots, J \tag{2.12}$$

Let's assume k*th* initial guesses $(x_{k1}, \dots, x_{kJ})$ matches the true toxicity curve and pr($M_k$) is the true model probability where $M_k$ is the true model. If there is no preference to have a priority for any single model in the CRM, then we can take $\text{pr}(M_k) = \frac{1}{K}$. Otherwise, a higher prior model probability can assign. Under this conditions, if we have a data,$D = \{(n_j, \dots, y_j), \ j = 1, \dots, J\}$, the likelihood function is,

$$L(D\backslash\alpha_k, M_k) = \prod_{j=1}^{J} \{{x_{kj}}^{\exp(\alpha_k)}\}^{y_j} \{1 - {x_{kj}}^{\exp(\alpha_k)}\}^{n_j - y_j} \tag{2.13}$$

So the posterior model probability of $M_k$,

$$\text{pr}(M_k\backslash D) = \frac{L(D\backslash M_k)\text{pr}(M_k)}{\sum_{i=1}^{K} L(D\backslash M_i)\text{pr}(M_i)} \tag{2.14}$$

Here, $\{L(D\backslash M_k)\}$ is marginal likelihood of $M_k$ and,

$$L(D\backslash M_k) = \int L(D\backslash\alpha_k, M_k) f(\alpha_k\backslash M_k)\, d\alpha_k \tag{2.15}$$

Where $\alpha_k$ is the power parameter and $f(\alpha_k\backslash M_k)$ is the prior distribution of $\alpha_k$ under $M_k$. Because of the possibility of a relationship between the posterior model probability and the Bayes factor, we need to formulate the Bayes factor, $B_{10}$, for a model $M_1$ against $M_0$,

$$B_{10} = \frac{pr(D\backslash M_1)}{pr(D\backslash M_0)} \tag{2.16}$$

Here, $pr(D\backslash M_k)$ is the same with $L(D\backslash M_k)$. If this formulation is adapted in general form, where, $n_k = \frac{pr(D\backslash M_k)}{pr(D\backslash M_0)}$,

$$pr(M_k\backslash D) = \frac{n_k B_{k0}}{\sum_{i=1}^{K} n_k B_{i0}} \tag{2.17}$$

The toxicity probability at each dose level,

$$\bar{\psi}_j = \sum_{i=1}^{K} \tilde{\psi}_{kj}\, pr(M_k\backslash D), \qquad j = 1, \dots, J \tag{2.18}$$

The posterior mean of toxicity probability,

$$\tilde{\psi}_{kj} = \int {x_{kj}}^{\exp(\alpha_k)} \frac{L(D\backslash \alpha_k, M_k) f(\alpha_k\backslash M_k)}{\int_0^{\infty} L(D\backslash \alpha_k, M_k) f(\alpha_k\backslash M_k) d\alpha_k} d\alpha_k \tag{2.19}$$

(2.19) identifies the best-fitting model. Escalation or de-escalation is made based on $\bar{\psi}_j$. Moreover, this method considers several sets of skeletons and updates the posterior model probabilities for all se of $x_{kj}$'s [58]. Dose-finding algorithm for the BMA-CRM is same with the CRM dose-finding algorithm.

## 2.8 Bayesian Optimal Interval Designs

The 3+3 design is the most dominant trial based on its simplicity. The BOIN design is similarly easy to implement and it is also flexible for choosing the target toxicity rate according to [69]. This method minimizes the probability of inappropriate dose assignments for the patients [70]. So, this design has a lower risk of overdosing on the patients.

In this design, escalation and de-escalation boundaries are pre-specified. A simple comparison of the observed DLT and pre-specified boundaries determines the dose escalation and de-escalation [69]. Simply, let $\hat{x}$ denote the observed DLT at the current dose level,

$$\hat{x} = \frac{patients\ who\ \ experienced\ DLT\ at\ the\ current\ dose\ level}{all\ patients\ treated\ at\ the\ current\ dose\ level} \quad (2.20)$$

Suppose $\lambda_e$ and $\lambda_d$ denote pre-specified boundaries, escalation and, de-escalation respectively. The algorithm of the BOIN design described as follows,

1. Start at the lowest dose level
2. Treat a patient or cohort of patients.
3. If it reaches the maximum sample size, stop the trial and select the MTD. Otherwise, compute the DLT rate at the current dose.
    - If DLT $\leq \lambda_{1i}$ ($\lambda_{1i}$ is the boundary) escalate
    - If DLT$\geq \lambda_{2j}$ deescalate
    - $\lambda_{1i} \leq$ DLT $\leq \lambda_{2j}$ retain
4. Back to step 2 until maximum sample size is reached.

Let $x_j$ denote the true DLT at the current dose level j. The theory under the BOIN design required three-point hypothesis:

- $H_1: x_j = \phi$ , Current dose is the MTD
- $H_2: x_j = \phi_1$, Current dose is below the MTD
- $H_3: x_j = \phi_2$, Current dose is above the MTD

where $\phi_1$ represents the highest toxicity probability below the MTD. In such cases, dose escalation need to be made. On the other hand, $\phi_2$ denotes the lowest toxicity that is overly toxic and in such cases, de-escalation is required.

To calculate the expected decision error and appropriate boundaries, equal prior probabilities can be assigned to each hypothesis under Bayesian approach. The boundaries $\lambda_e$ and $\lambda_d$ can be calculated as,

$$\lambda_e = \frac{\log(1-\phi_1/1-\phi)}{log\ (\phi(1-\phi_1)/\phi_1(1-\phi))} \quad (2.21)$$

And,

$$\lambda_d = \frac{\log(1-\phi/1-\phi_2)}{log\ (\phi_2(1-\phi)/\phi(1-\phi_2))} \quad (2.22)$$

Remarkably, the BOIN design is independent from dose level and the number of patients [71]. Obtained boundaries can be used until the trial is completed. The flowchart of the BOIN design is shown in Figure 2.7.

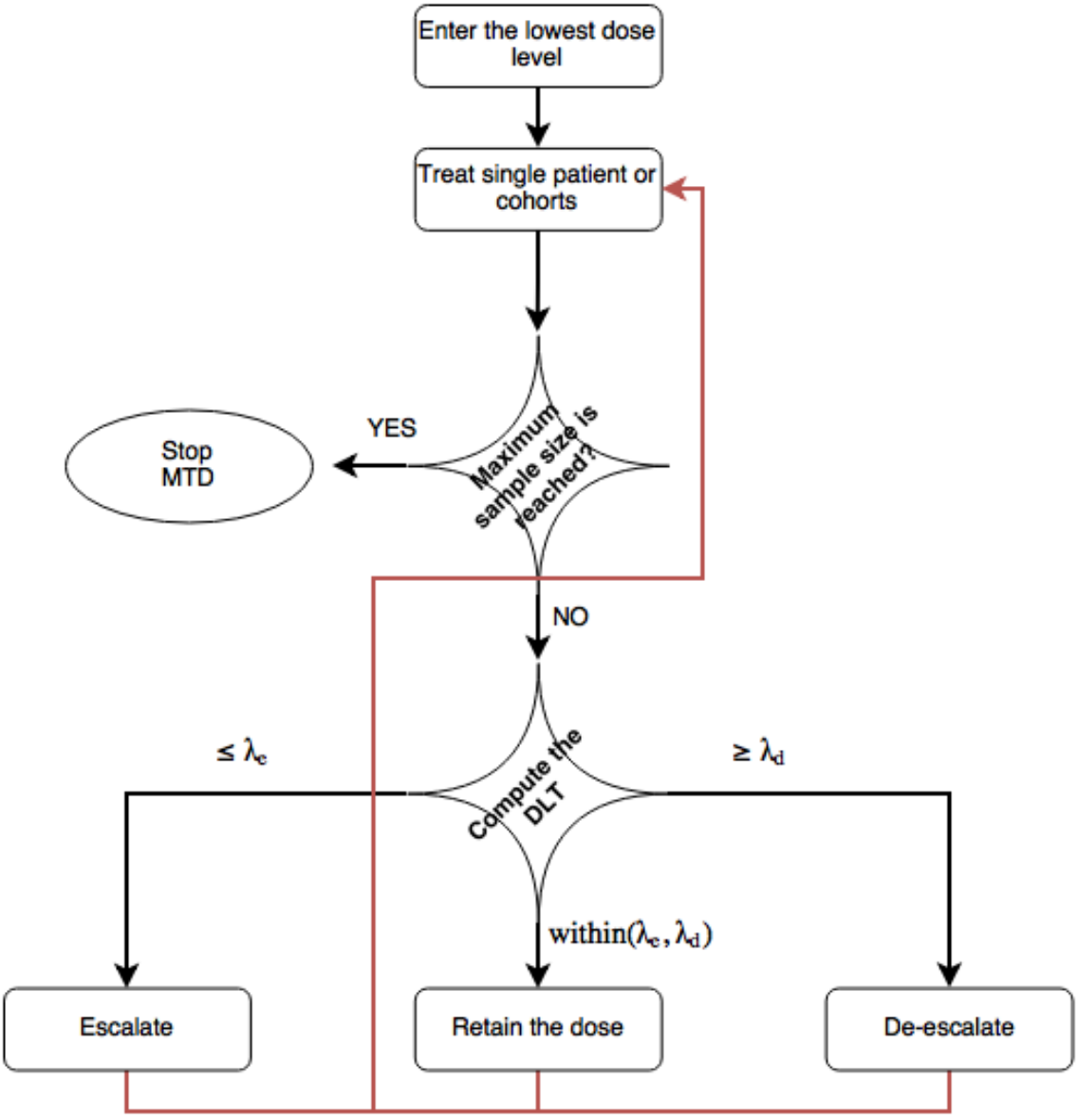


Figure 2.8 Flowchart of the BOIN

## 2.9 Modified Toxicity Probability Interval Method

A modified toxicity probability interval (mTPI) is a method that proposes dose-finding-decision rules based on the unit probability mass (UPM). This method is the improved version of the toxicity probability interval method (TPI), and it uses beta-binomial

hierarchical model [72]. The UPM is used to determine decision rules of three intervals corresponding to low, high and proper dosing in terms of toxicity [48]. Although the mTPI allows researchers to understand the decisions before the trials start, some decision rules are debated in practice [73].

This design is the extended version of the TPI design. In the TPI design, the performance of the result depends on the two key parameters, $K_1, K_2$. These parameters define the toxicity intervals. Different choice of $K_1, K_2$ could lead to different results [72]. Although, the mTPI design is similar to the TPI design [48], the mTPI method does not have to calibrate the design for different trials. So that this design is calibration free [72]. Secondly, $K_1$ and $K_2$ are used in the posterior probabilities of the 3 different intervals. However, the mTPI design pre-specified the equivalence intervals (IE) before the trial. In addition, the intervals does not depend on any parameters of the probability model [72]. Suppose we have an EI,$[\, p_t - \varepsilon_1, p_t + \varepsilon_2]$, where $\varepsilon_1$ and $\varepsilon_2$ are small fractions and they help to account for uncertainty around the true target toxicity. In addition, $p_i$ denotes the unknown probability of the toxicity at $d$ dose levels, $i = 1, \dots, d$. Let $n_i$ be the subjects treaded and $x_i$ is the subjects experienced the toxicity. After all subjects are treated in the trial, 3 outcomes can be obtained; escalate the dose level, E, $(i + 1)$; deescalate the dose level, *D*, $(i - 1)$; stay at the same dose level, *S*, i. The defined EI, $[p_t - \varepsilon_1, p_t + \varepsilon_2]$ should be determined after collaborating with clinicians. Then, three different intervals will be determined from the EI, which are $(0, p_t - \varepsilon_1)$, $[p_t - \varepsilon_1, p_t + \varepsilon_2]$, $(p_t + \varepsilon_2, 1)$. These intervals represent the toxicities; lower than the MTD, close to the MTD, higher than the MTD respectively. In the light of above information, the algorithm of the mTPI design is very simple. The algorithm is described below.

- Assume the dose levels in trial is $i, i\epsilon\{1, \dots, d\}$. After all toxicity results of the cohorts are observed, select one of the interval which has the highest UPM; $(0, p_t - \varepsilon_1)$, $[p_t - \varepsilon_1, p_t + \varepsilon_2]$, $(p_t + \varepsilon_2, 1)$.
- If $Pr(p_1 > p_t|data) > 0.95$, stop the trial.
- If the decision is escalation from dose *i* to *i+1* and $Pr(p_{i+1} > p_t|data) > 0.95$, treat the subject at dose i. Never use that dose level again.
- Let $\widehat{p_i}$ is a sensible estimate of $p_i$. Select the dose as the MTD, where $|\widehat{p_i} - p_t|$ is smallest. If more than one doses are found as the MTD, (where $\hat{p}^*$ is $\widehat{p_i}$ of the tied dose);

- $\hat{p}^* < p_t$ choose the highest one
- $\hat{p}^* > p_t$ choose the lowest dose

The detailed information behind the TPI and mTPI design can be read in [48], [72].

## 2.10 A Bayesian Interval Dose-Finding Design Addressing Ockham's Razor

In the mTPI design, some decision rules are debated in practice. To prevent such debates, Yang et al. [74] proposed an ad-hoc remedy that allows the decision rules in the mTPI design. However, it is lack of solid statistical justification and cannot be properly assessed. To handle this problem, mTPI-2 was developed. It solves the undesirable issue in the current decision under the mTPI [73]. This design blunts the Ockham's razor for the mTPI. Ockham's razor is a logical principle established by the logician William in the 14*th* century. This principle points that one should not make more assumption than the minimum needed [75], [76], [77]. The impact of this principle on mTPI is on interval selection.

To blunt the Ockham's razor for mTPI, the unit interval $(0,1)$ need to be divided into subintervals with equal length. Let EI be the interval $[\,p_t - \varepsilon_1, p_t + \varepsilon_2]$, we now call the set of intervals below the EI as LI and above the EI as HI. So if the EI, $M^{EI} = [\,p_t - \varepsilon_1, p_t + \varepsilon_2]$ has the largest UPM, it is selected as the MTD. If any other interval $M^{HI}$ or $M^{LI}$ has the largest UPM, the dose-finding decision is de-escalation or escalation, respectively. The optimal rule for mTPI-2 considers $(0,1)$ loss function. For more information on the theory of optimal decision rule and $D_{mTPI-2}$ see [73]. The design algorithm of the mTPI-2 is similar to mTPI and described below.

- Assume the current dose level is $d, d\epsilon\{1, \dots, D\}$. After the toxicity results of the last cohort of subjects are observed, define $(x_d, n_d)$ the current observed data. Select the dose for the next group of subject $\{(d-1), d, (d+1)\}$ based on the optimal rule $D_{mTPI-2}$.
- Suppose $n_1 > 0$. If $Pr(p_1 > p_t | x_d, n_d) > 0.95$, stop the trial. Stopping rule is not applied for the case of cohort size of one.
- Let $\widehat{p_i}$ is a sensible estimate of $p_i$. Select the dose as the MTD, where $|\widehat{p_i} - p_t|$ is smallest. If more than one doses are found as the MTD, (where $\hat{p}^*$ is $\widehat{p_i}$ of the tied dose);
    - $\hat{p}^* < p_t$ choose the highest one

- $\hat{p}^* > p_t$ choose the lowest dose

Remarkably, this design is able to show Bayes factors for each decision. This may help clinicians in order to compare the two decisions which are close to one. Table 2.1 summarizes the algorithms and limitations of the introduced methods in Chapter 2.

Table 2.2 Summary of Selected Phase 1 Dose Escalation Methods

| Method | Setup | Limitations |
|---|---|---|
| **3+3 Design** | e) Assess the toxicity of 3 treated patients at jth dose level<br>f) If any of the 3 patients' dose does not reach DLT, the dose level is increased to j+1 and then back to step 1<br>g) If one of three patients encounter DLT, 3 patients are treated with the same dose level j and the process is as follows:<br>• If one of six patients encounters DLT, the j dose level is increased to the dose level j+1 if j is non-exceeding MTD<br>• If two out of six patients meet the DLT, the experiment is terminated and a sub-six dose level is identified as j-1 MTD.<br>• If more than 2 patients meet with DLT, the current dose level is considered as exceeded the MTD and 3 patients will be treated at j-1 level and less than 6 patients will be treated at this dose level.<br>h) If two or three patients meet DLT (dose level exceeded MTD), 3 patients will be treated at j-1 dose level, based on the fact that less than 6 patients are treated at j-1 dose level. | 1. The only previous dose is included. Other dosage history is ignored.<br>2. It ignores uncertainty<br>3. Most of the patients are treated at a low dose level.<br>4. Selecting MTD is with low probability.<br>5. Cohort sizes are 3 or 6.<br>6. MTD with target probability of DLT < 20% or >33% cannot be estimated. |
| **CRM** | 1. Start with a prior estimate of DLT for each dose level.<br>2. Select a mathematical model to describe the relationship between dose and DLT.<br>3. Describe uncertainty about the model by prior distribution.<br>4. After each patient, update the model, and estimate the probability of toxicity at each dose level.<br>5. Treat the next patient at the dose whose estimate is the closest to some pre-specified target.<br>6. Stop when a maximum sample size is reached. | 1. Selection of skeleton is very important for MTD.<br>2. Sometimes MTD is overestimated. |
| **BOIN** | 5. Start at the lowest dose level<br>6. Treat a patient or cohort of patients.<br>7. If it reaches the maximum sample size, stop the trial and select the MTD. Otherwise, compute the DLT rate at the current dose.<br>• If DLT $\leq \lambda_{1i}$ ($\lambda_{1i}$ is the boundary) escalate<br>• If DLT$\geq \lambda_{2j}$ deescalate<br>• $\lambda_{1i} \leq$ DLT $\leq \lambda_{2j}$ retain<br>8. Back to step 2 until maximum sample size is reached. | 1. Results can be unreliable due to small sample size.<br>2. The inconclusive field does not change even if the information is accumulated during the trial. |
| **mTPI** | 1. The probability of toxicity at each dose modeled by beta distributions.<br>2. Set of decision intervals specified.<br>3. Dosing decisions determined by normalized posterior probability in each interval at the current dose $d_i$:<br>• Escalate to $d_{i+1}$ if $d_i$ is underdosing<br>• Stay at $d_i$ if $d_i$ is proper<br>• De-escalate $d_{i-1}$ if $d_i$ is overdosing<br>4. Compute UPMs and the largest one implies the decision. | 1. Some patients are treated at doses over MTD.<br>2. Applicable to trials with a binary toxicity endpoint. |
| **mTPI-2** | Same procedures are followed as in mTPI. However, probability models are sharpened by the Ocham's razor problem is blunted. | 1. Sample size depends on the number of doses. |
| **BMA-CRM** | 1. Obtain the probability that satisfies minimum efficacy.<br>2. Calculate the Euclidean distance by using BMA<br>3. The recommended dose is determined when the maximum number of patients is reached based on all accumulated outcomes. | 4. The selection of skeleton is very important for MTD. |
| **BCRM** | 1. Assign the dose level for the first cohort.<br>2. Terminate the trial when the first dose is shown to be toxic.<br>3. Prevent treating patients at toxic doses.<br>4. Determine the most appropriate dose for the next cohort of patients. | 1. The dose assignment maybe too aggressive.<br>2. Needs a statistical software to implement the design |

# CHAPTER 3

# INTRODUCTION TO REAL LIFE STORIES

This chapter introduces the three different stories of the simulation studies. The stories used in the study were compiled from finished and last-run projects. The specified disease name, characteristics, dose levels, and dose rates were included in the study without modification. The stories introduced in the scenario I and scenario II are used to compare the most commonly used methods in the literature in the first part of the study. In the scenario III, the highlighted methods in the scenario I and II were compared with the Bayesian CRM when it has different model structures and priors. In addition, a simulation analysis performed using bcrm, BOIN, dfcrm, TERAplusB and UBCRM packages in R Statistical software.

## 3.1 Scenario I

In the first story, we illustrated the proposed designs using a pharmacokinetic phase I clinical trial that aimed to investigate the safety/tolerability and the pharmacokinetic profile of SHR6390 (small molecular, oral potent, selective CDK4/6) in Chinese advanced melanoma patients. Each subject receives a single dose of SHR6390 and then repeats doses following a 3 week/1 week off regimen. The age range of the patients is between 18 years to 65 years old. All sexes are included in the study. This trial studied six different dose levels of SHR6390: 50 mg, 75 mg, 100 mg, 125 mg, 150 mg and 175 mg [78]. The number of patients included in the study was 30. We considered the MTD as the dose with a DLT rate of 30% and elicited five different skeletons.

$$(p_1, p_2, p_3, p_4, p_5, p_6) = \begin{cases} (0.18, 0.30, 0.42, 0.53, 0.64, 0.72) & skeleton\ 1 \\ (0.02, 0.07, 0.16, 0.30, 0.44, 0.57) & skeleton\ 2 \\ (0.05, 0.15, 0.30, 0.46\ 0.61, 0.73) & skeleton\ 3 \\ (0.02, 0.06, 0.12, 0.20, 0.30, 0.40) & skeleton\ 4 \\ (0.20, 0.22, 0.24, 0.26, 0.28, 0.30) & skeleton\ 5 \end{cases}$$

## 3.2 Scenario II

In the second story, we applied seven designs to a phase I Reverse Genetic reassortant H9N2 influenza vaccine study conducted at Nanotherapeutics, Inc. The aim of the study was to identify the optimal dose level of a reverse genetic reassortant H9N2 influenza vaccine for further product development. The age range of the patients is between 18-65 years old. All sexes are included in the study. This clinical trial studied six different dose levels of a Reverse Genetic reassortant H9N2 pandemic influenza vaccine in healthy subjects aged 18 to 49 Years: 3.75 µg, 7.5 µg, 15 µg, 30 µg, 45 µg or 60 µg [79]. The number of the patients included in the study is 21. We considered the MTD as the dose with a DLT rate of 25% and elicited four different skeletons,

$$(p_1, p_2, p_3, p_4, p_5, p_6) = \begin{cases} (0.07, 0.14, 0.25, 0.37, 0.50, 0.61) & skeleton\ 1 \\ (0.06, 0.11, 0.17, 0.25, 0.33, 0.39) & skeleton\ 2 \\ (0.17, 0.21, 0.25, 0.29, 0.33, 0.37) & skeleton\ 3 \\ (0.20, 0.25, 0.44, 0.61, 0.75, 0.84) & skeleton\ 4 \end{cases}$$

## 3.3 Scenario III

In the third story, we implemented the dose-escalating trial of Tanibirumab in our simulation study. Tanibirumab is a human monoclonal antibody to vascular endothelial growth factor receptor 2 (VEGFR2/KDR). This study enrolls patients with metastatic cancer or the last phase cancer who are refractory or for patients who do not have standard therapeutic options.

Each patient receives a Tanibirumab intravenously over 60 minutes on day 1, 8 and 15. The length of the treatment is a minimum of 28 days. The dose finding methods are designed to identify the RP2D which will be based on safety, tolerability, and pharmacokinetics of the RP2D. The eligible age of the patients for the study is over 20 years old. All sexes are included in the study. This trial studied six different dose levels to be potentially tested in the phase I include; 1 mg, 2 mg, 4 mg, 8 mg, 12 mg and 16 mg. The total dose of Tanibirumab for each subject depends on dose level assignment on the subject's weight [80].

The number of the patients included in the study was 30. In this simulation, we considered the MTD as the dose with a DLT rate of 30% and elicited five different skeletons.

$$(p_1, p_2, p_3, p_4, p_5, p_6) = \begin{cases} (0.14, 0.19, 0.24, 0.30, 0.36, 0.42) & skeleton\ 1 \\ (0.01, 0.06, 0.30, 0.60, 0.78, 0.86) & skeleton\ 2 \\ (0.06, 0.10, 0.16, 0.22\ 0.30, 0.38) & skeleton\ 3 \\ (0.30, 0.50, 0.66, 0.77, 0.83, 0.87) & skeleton\ 4 \\ (0.05, 0.30, 0.60, 0.78, 0.86, 0.90) & skeleton\ 5 \end{cases}$$

# CHAPTER 4

# SIMULATION STUDY

The drug development is a risky and complex process that many drug candidates are failing. The attrition rate challenge in pharmaceutical research and development can be reduced by using the simulation studies in the clinical trials. The simulation studies help to predict the outcomes of the clinical trials by exploring how different trial designs will perform to detect the drug effects and characteristics. Moreover, it shows the impact of differences in dosing regimen patient profile, sample size, trial duration and choice of comparators, so that the most effective design can be picked after simulation study. In addition, during the early development stage, simulation helps to estimate the likelihood of meeting the efficacy phase to criteria. The design can be optimized and searched for ways to make it more cost-effective. In this chapter, simulation studies for the scenario I and the scenario II are used to compare the performance of the most commonly used methods in the literature. Furthermore, the simulation study for the scenario III is used to compare the highlighted methods in the scenario I-II and Bayesian CRM when it has different model structures and priors. All simulation analysis performed using R statistical software.

## 4.1 Simulation Study for Scenario (Story) I

The scenario I is used in the first simulation run. We applied introduced seven designs in this simulation study. It is important that the BCRM is the design that is built with the model that gives the best performance when different model structures are taken in the CRM. In the BMA-CRM, CRM, and BCRM, the skeletons were very important in the selection of the MTD. This is because skeletons represent different prior opinions and it leads to produce different MTD selections. Skeletons represent different prior guesses of the toxicity profile of the drug. The relation between true toxicity rate and prior probabilities is very important because selection probability of the MTD may increase if

the prior probabilities are similar to the true toxicity probabilities. For the CRM and the BCRM models, all skeletons were assigned one by one to the models and the best-performed one was used for comparison. The first skeleton is for the case where toxicity starts at a high level and increases with the almost same rate. In the second skeleton, toxicity increases slowly when the dose is low but increases quickly at the high doses. The third skeleton starts with low dose level and increases with high dose levels.

The toxicity probabilities in the fourth skeleton are more gathered at the low toxicity levels. The last skeleton is concentrated in a narrow range where the toxicity probability starts at 0.2 and ends at 0.3. Table 4.1-4.6 shows the simulation results of the scenarios (simulation scenarios) for the 3+3 design, CRM, BMA-CRM, BCRM, mTPI, mTPI-2, and BOIN. The probability selection of the MTD and the number of treated patients are given in the tables. We carried out 10,000 simulations for each scenario.

Table 4.1 Comparison of scenario I with a toxicity target 30%

| | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/d** | **75mg/d** | **100mg/d** | **125mg/d** | **150mg/d** | **175mg/d** |
| | **Method** | True toxicity rate | 0.02 | 0.06 | 0.12 | 0.2 | **0.3** | 0.4 |
| **Scenario1** | **3+3** | Probability selection | 0.039 | 0.153 | 0.235 | 0.325 | **0.174** | 0.074 |
| | | #of patient treated | 3.243 | 3.87 | 4.296 | 4.086 | **3.036** | 1.215 |
| | **CRM** | Probability selection | 0 | 0 | 0.032 | 0.277 | **0.468** | 0.223 |
| | | #of patient treated | 3.24 | 3.204 | 4.422 | 7.656 | **7.383** | 4.095 |
| | **BMA-CRM** | Probability selection | 0 | 0 | 0.03 | 0.28 | **0.421** | 0.27 |
| | | #of patient treated | 3.3 | 3.7 | 5 | 7.6 | **6.8** | 3.6 |
| | **BCRM** | Probability selection | 0 | 0.002 | 0.035 | 0.334 | **0.445** | 0.184 |
| | | #of patient treated | 3.156 | 3.1 | 4.342 | 6.004 | **7.18** | 3.987 |
| | **mTPI** | Probability selection | 0.001 | 0.006 | 0.079 | 0.318 | **0.372** | 0.224 |
| | | #of patient treated | 3.267 | 4.074 | 5.895 | 7.662 | **6.015** | 2.087 |
| | **mTPI-2** | Probability selection | 0.001 | 0.006 | 0.061 | 0.287 | **0.393** | 0.252 |
| | | #of patient treated | 3.261 | 4.017 | 5.64 | 7.485 | **6.24** | 3.357 |
| | **BOIN** | Probability selection | 0 | 0.007 | 0.07 | 0.307 | **0.383** | 0.233 |
| | | #of patient treated | 3.2 | 4.1 | 5.8 | 7.6 | **6.1** | 3.2 |

The A+B design was tested during the simulation runs. However, different variations of the A+B design produced very poor selection probabilities in comparison to the 3+3 design. So the results of other variations of the A+B design were not included in the tables. In Appendix B, the tables show the selection probabilities of different variations of the A+B design. In the first scenario, the fifth dose was the MTD. However, the 3+3 design had the lowest selection percentage of 17.4% and selected the fourth dose as the MTD with a percentage of 32.4%. On the other hand, the BMA-CRM selected the MTD with 42.1% and selection of the BCRM performed slightly better than the BMA-CRM with 44.5%. The CRM performed better than the other methods for the selection of the true MTD with 46.8%.

Table 4.2 Comparison of scenario II with a toxicity target 30%

| | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/d** | **75mg/d** | **100mg/d** | **125mg/d** | **150mg/d** | **175mg/d** |
| | **Method** | True toxicity rate | 0.05 | 0.15 | **0.3** | 0.46 | 0.61 | 0.73 |
| **Scenario2** | **3+3** | Probability selection | 0.217 | 0.268 | **0.428** | 0.057 | 0.030 | 0 |
| | | #of patient treated | 3.966 | 4.147 | **5.978** | 1.776 | 0.375 | 0.018 |
| | **CRM** | Probability selection | 0.002 | 0.169 | **0.619** | 0.204 | 0.005 | 0.001 |
| | | #of patient treated | 4.17 | 7.152 | **11.82** | 5.901 | 0.906 | 0.051 |
| | **BMA-CRM** | Probability selection | 0 | 0.16 | **0.6** | 0.24 | 0.001 | 0 |
| | | #of patient treated | 4.1 | 7.4 | **12.1** | 5.9 | 0.6 | 0.9 |
| | **BCRM** | Probability selection | 0.006 | 0.161 | **0.577** | 0.249 | 0.007 | 0 |
| | | #of patient treated | 3.825 | 6.707 | **11.12** | 5.611 | 0.94 | 0.057 |
| | **mTPI** | Probability selection | 0.015 | 0.215 | **0.549** | 0.206 | 0.015 | 0 |
| | | #of patient treated | 4.182 | 8.856 | **11.76** | 4.509 | 0.657 | 0.036 |
| | **mTPI-2** | Probability selection | 0.015 | 0.219 | **0.548** | 0.2 | 0.018 | 0 |
| | | #of patient treated | 4.284 | 9.159 | **11.33** | 4.569 | 0.819 | 0.036 |
| | **BOIN** | Probability selection | 0.001 | 0.235 | **0.559** | 0.182 | 0.001 | 0 |
| | | #of patient treated | 4.2 | 9.3 | **11.1** | 4.7 | 0.7 | 0 |

In the second scenario, the third dose was the MTD. The worst selection of the MTD was made by 3+3 design with 42.8% and the MTD selection percentage using the CRM

was the best among the other designs. The BMA-CRM was the second best among the others.

Table 4.3 Comparison of scenario III with a toxicity target 30%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/d** | **75 mg/d** | **100mg/d** | **125 mg/d** | **150mg/d** | **175mg/d** |
| | **Method** | True toxicity rate | 0.02 | 0.07 | 0.16 | **0.3** | 0.44 | 0.57 |
| **Scenario3** | **3+3** | Probability selection | 0.054 | 0.215 | 0.265 | **0.395** | 0.06 | 0.005 |
| | | #of patient treated | 3.33 | 4.131 | 4.683 | **5.846** | 1.764 | 0.333 |
| | **CRM** | Probability selection | 0 | 0.004 | 0.167 | **0.583** | 0.231 | 0.015 |
| | | #of patient treated | 3.276 | 3.474 | 6.741 | **10.37** | 5.118 | 1.014 |
| | **BMA-CRM** | Probability selection | 0 | 0 | 0.17 | **0.562** | 0.25 | 0.03 |
| | | #of patient treated | 3.3 | 3.9 | 7.1 | **10.2** | 4.7 | 0.8 |
| | **BCRM** | Probability selection | 0 | 0.007 | 0.177 | **0.58** | 0.224 | 0.012 |
| | | #of patient treated | 3.154 | 3.2 | 6.245 | **10.01** | 5.111 | 1.007 |
| | **mTPI** | Probability selection | 0.008 | 0.016 | 0.237 | **0.53** | 0.194 | 0.022 |
| | | #of patient treated | 3.297 | 4.488 | 8.328 | **9.507** | 3.762 | 0.618 |
| | **mTPI-2** | Probability selection | 0.014 | 0.025 | 0.223 | **0.536** | 0.195 | 0.003 |
| | | #of patient treated | 3.303 | 4.518 | 8.247 | **9.279** | 3.918 | 0.735 |
| | **BOIN** | Probability selection | 0.001 | 0.017 | 0.246 | **0.49** | 0.22 | 0.026 |
| | | #of patient treated | 3.3 | 4.6 | 8.2 | **9.2** | 4 | 0.7 |

Scenario 3 had the MTD at the fourth dose level, and the MTD selection using the CRM and BCRM performed the best results with almost similar percentages. The BMA-CRM and mTPI-2 performed well, with the MTD selection probabilities of 56.2% and 53.6% respectively and the mTPI produced almost similar MTD selection probability with the mTPI-2.

Table 4.4 Comparison of scenario IV with a toxicity target 30%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/d** | **75 mg/d** | **100mg/d** | **125 mg/d** | **150mg/d** | **175mg/d** |
| | **Method** | True toxicity rate | 0.05 | 0.1 | 0.15 | 0.2 | **0.3** | 0.4 |
| **Scenario4** | **3+3** | Probability selection | 0.105 | 0.165 | 0.208 | 0.253 | **0.297** | 0.078 |
| | | #of patient treated | 3.636 | 4.131 | 3.987 | 3.495 | **4.595** | 1.058 |
| | **CRM** | Probability selection | 0 | 0.013 | 0.082 | 0.308 | **0.427** | 0.17 |
| | | #of patient treated | 3.735 | 3.807 | 5.346 | 7.683 | **6.432** | 2.997 |
| | **BMA-CRM** | Probability selection | 0 | 0.01 | 0.1 | 0.32 | **0.411** | 0.166 |
| | | #of patient treated | 8.1 | 7.3 | 5.9 | 3.8 | **2** | 0.9 |
| | **BCRM** | Probability selection | 0 | 0.008 | 0.049 | 0.267 | **0.589** | 0.287 |
| | | #of patient treated | 3.7 | 4.605 | 6.102 | 7.51 | **6.022** | 2.327 |
| | **mTPI** | Probability selection | 0.003 | 0.028 | 0.031 | 0.209 | **0.532** | 0.197 |
| | | #of patient treated | 3.834 | 5.151 | 6.528 | 6.873 | **4.992** | 2.622 |
| | **mTPI-2** | Probability selection | 0.003 | 0.023 | 0.096 | 0.19 | **0.558** | 0.13 |
| | | #of patient treated | 4.284 | 9.159 | 11.33 | 4.569 | **0.819** | 0.036 |
| | **BOIN** | Probability selection | 0.002 | 0.018 | 0.103 | 0.302 | **0.365** | 0.21 |
| | | #of patient treated | 3.804 | 4.872 | 6.096 | 6.906 | **5.394** | 2.928 |

In scenario 4, the BCRM and mTPI-2 produced the best MTD selection probabilities of 58.9% and 55.8%. Interestingly, The BMA-CRM and the CRM did not perform well in this scenario. The worst selection of the MTD was made by the 3+3 design with 29.7%. In this scenario, the selection of the model structure was remarkably effected on the selection of the true MTD.

Table 4.5 Comparison of scenario V with a toxicity target 30%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/d** | **75 mg/d** | **100mg/d** | **125mg/d** | **150mg/d** | **175mg/d** |
| | **Method** | True toxicity rate | 0.18 | **0.3** | 0.42 | 0.53 | 0.64 | 0.72 |
| **Scenario5** | **3+3** | Probability selection | 0.311 | **0.425** | 0.171 | 0.095 | 0 | 0 |
| | | #of patient treated | 5.049 | **3.795** | 1.782 | 0.453 | 0.06 | 0.006 |
| | **CRM** | Probability selection | 0.28 | **0.529** | 0.206 | 0.039 | 0 | 0 |
| | | #of patient treated | 10.92 | **11.05** | 6.147 | 1.332 | 0.156 | 0.006 |
| | **BMA-CRM** | Probability selection | 0.32 | **0.56** | 0.12 | 0 | 0 | 0 |
| | | #of patient treated | 10.6 | **12** | 5.8 | 1 | 0.1 | 0 |
| | **BCRM** | Probability selection | 0.17 | **0.571** | 0.242 | 0.017 | 0 | 0 |
| | | #of patient treated | 10.15 | **11.8** | 6.24 | 1.101 | 0.122 | 0.007 |
| | **mTPI** | Probability selection | 0.245 | **0.508** | 0.202 | 0.023 | 0.002 | 0 |
| | | #of patient treated | 10.93 | **12.5** | 5.214 | 0.9 | 0.069 | 0.003 |
| | **mTPI-2** | Probability selection | 0.231 | **0.524** | 0.205 | 0.028 | 0.003 | 0 |
| | | #of patient treated | 11.07 | **11.9** | 5.4 | 1.164 | 0.078 | 0.003 |
| | **BOIN** | Probability selection | 0.254 | **0.512** | 0.18 | 0.048 | 0.004 | 0 |
| | | #of patient treated | 10.9 | **12.1** | 5.3 | 1.1 | 0.2 | 0 |

In scenario 5, the BCRM design showed the best performance and the BMA-CRM design has the second best performance. Similarly, the 3+3 design has the worst performance. The other designs have similar performances. However, the probability of selecting the previous dose as MTD was very similar to all the other designs.

Table 4.6 Comparison of scenario VI with a toxicity target 30%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/d** | **75 mg/d** | **100mg/d** | **125mg/d** | **150mg/d** | **175mg/d** |
| | **Method** | True toxicity rate | 0.01 | **0.3** | 0.55 | 0.65 | 0.8 | 0.95 |
| **Scenario6** | **3+3** | Probability selection | 0.345 | **0.465** | 0.168 | 0.003 | 0 | 0 |
| | | #of patient treated | 3.636 | **4.131** | 3.987 | 3.495 | 2.595 | 1.058 |
| | **CRM** | Probability selection | 0.074 | **0.767** | 0.158 | 0.001 | 0 | 0 |
| | | #of patient treated | 6.312 | **15.74** | 7.047 | 0.837 | 0.066 | 0 |
| | **BMA-CRM** | Probability selection | 0.05 | **0.8** | 0.15 | 0 | 0 | 0 |
| | | #of patient treated | 5.6 | **17.2** | 6.7 | 0.5 | 0 | 0 |
| | **BCRM** | Probability selection | 0.044 | **0.815** | 0.138 | 0.003 | 0 | 0 |
| | | #of patient treated | 6.126 | **16.62** | 6.917 | 0.919 | 0.055 | 0 |
| | **mTPI** | Probability selection | 0.142 | **0.765** | 0.09 | 0.003 | 0 | 0 |
| | | #of patient treated | 6.204 | **18.46** | 4.947 | 0.363 | 0.018 | 0 |
| | **mTPI-2** | Probability selection | 0.124 | **0.786** | 0.087 | 0.003 | 0 | 0 |
| | | #of patient treated | 7.998 | **16.63** | 4.995 | 0.363 | 0.018 | 0 |
| | **BOIN** | Probability selection | 0.164 | **0.759** | 0.073 | 0.004 | 0 | 0 |
| | | #of patient treated | 8 | **16.6** | 4.9 | 0.4 | 0 | 0 |

Scenario 6 had the MTD at the second dose level. All designs selected the true MTD correctly with very high selection probability rate. Although the 3+3 was selected the correct dose level as the MTD, the selection probability was low. In this scenario, the BCRM has the highest probability rate with 81.5% and the lowest probability rate was 46.5% which was produced by the 3+3 design.

All the other designs have similar selection probabilities. From these findings, it can be said that the MTD selection percentage of the first scenario, where the dose levels start at low then increase gradually to medium dose level, is not high for all designs. The selection of the skeleton is very important for the CRM and the BCRM. These designs can perform with the lowest selection percentage if the choice of the skeleton is not appropriate. The BMA-CRM and mTPI-2 performed very similarly to the CRM and the BCRM. The 3+3 design is very simple and easy to implement. However, this design had the lowest selection percentage for all scenarios.

The MTD selections of the designs can be seen clearly with histogram plot. Figure 4.1 shows the selection probabilities of the seven designs for the scenario I and the scenario II.

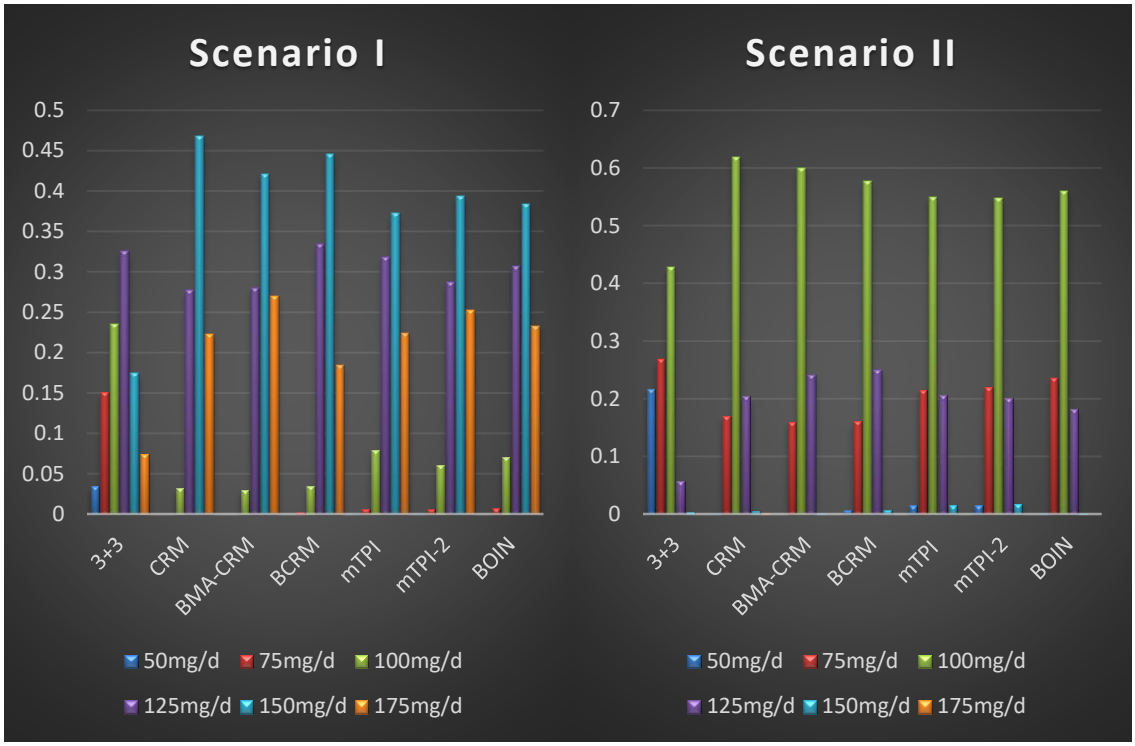


Figure 4.9 Comparison of the designs for scenario I and scenario II

Some histogram bars did not appear in the plot. This is because the selection probability of the dose level related to that histogram bar is zero. In the scenario I, it is clear that 3+3 design was selected the wrong dose level as the MTD. However, other designs were selected the MTD as $5^{th}$ dose level. Figure 4.2 shows the MTD selection of the seven designs for scenario III and scenario IV.

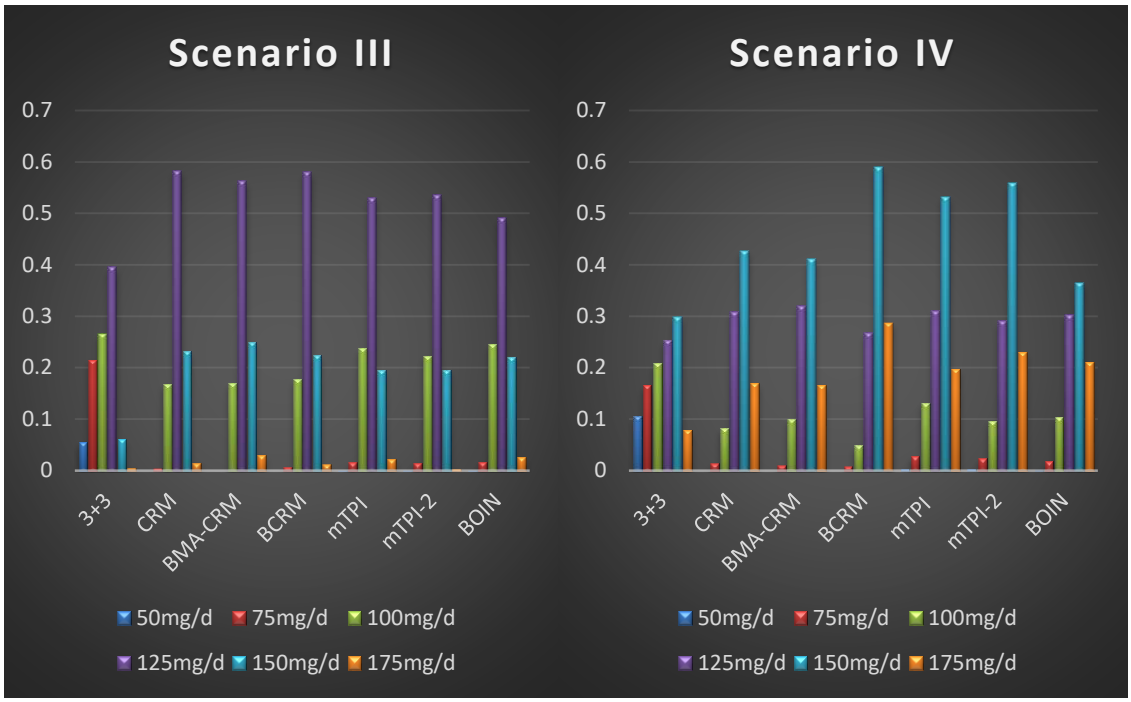


Figure 4.10 Comparison of the designs for scenario III and scenario IV

Observation of the selection probabilities can be difficult from the tables. However, histogram plots are very useful to illustrate the selection of the true dose levels. The Figure 4.3 shows the MTD selection of the seven designs for scenario V and scenario VI.

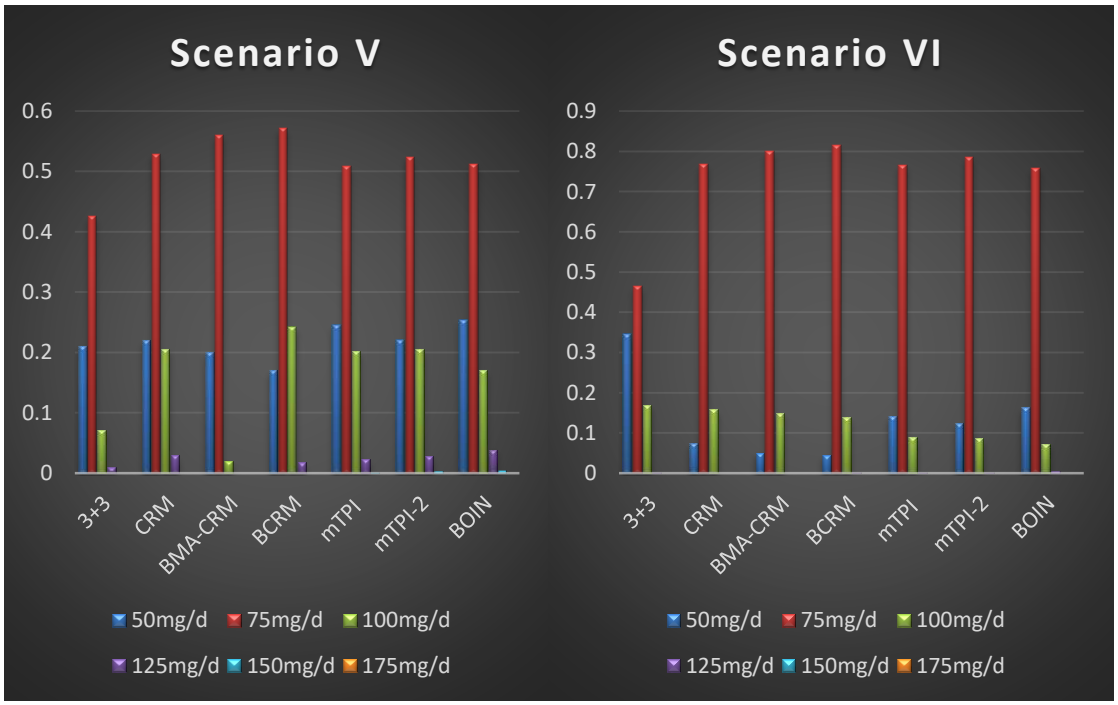


Figure 4.11 Comparison of the designs for scenario V and scenario VI

## 4.2 Simulation Study for Scenario (Story) II

The process of using the skeletons which are applied in the first story is repeated for the second story. Since the skeletons represent different prior opinions, it is important to choose the right skeleton for the CRM and the BCRM. The first skeleton starts at the low toxicity level then increases synchronously. The toxicity probabilities in the second skeleton are changed between more 0.06 and 0.39 (low toxicity levels). In the third skeleton, the toxicity increases slowly at all dose levels. The fourth skeleton starts with at a low dose level but increases quickly at the high doses. The simulation results of the second illustration for seven different designs (the 3+3 design, CRM, BMA-CRM, BCRM, mTPI, mTPI-2 and BOIN) are given in the Table 4.7-4.12. We list the true toxicity probability in the first row. The probability selection of the MTD and the numbers of the treated patients are given in the tables. However, the comparison is made with the probability selection of the MTD. The target toxicity probability was 25%. The sample size was 21 and 10,000 simulations are carried out for each scenario.

Table 4.7 Comparison of scenario I with a toxicity target 25%

| | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | 0.15 | 0.2 | **0.25** | 0.3 | 0.35 | 0.4 |
| **Scenario1** | **3+3** | Probability selection | 0.133 | 0. 165 | **0.283** | 0.203 | 0.134 | 0.117 |
| | | #of patient treated | 4.611 | 4.921 | **5.73** | 1.662 | 0.816 | 0.24 |
| | **CRM** | Probability selection | 0.15 | 0.304 | **0.318** | 0.158 | 0.05 | 0.011 |
| | | #of patient treated | 9.672 | 9.225 | **6.432** | 3.132 | 1.08 | 0.303 |
| | **BMA-CRM** | Probability selection | 0.11 | 0.231 | **0.28** | 0.172 | 0.101 | 0.066 |
| | | #of patient treated | 8.1 | 7.3 | **5.9** | 3.8 | 2 | 0.9 |
| | **BCRM** | Probability selection | 0.122 | 0.259 | **0.32** | 0.167 | 0.067 | 0.065 |
| | | #of patient treated | 9.245 | 8.977 | **6.511** | 3.444 | 1.214 | 0.877 |
| | **mTPI** | Probability selection | 0.198 | 0.327 | **0.263** | 0.136 | 0.041 | 0.016 |
| | | #of patient treated | 1.615 | 1.889 | **1.478** | 0.788 | 0.225 | 0.091 |
| | **mTPI-2** | Probability selection | 0.173 | 0.336 | **0.273** | 0.128 | 0.076 | 0.015 |
| | | #of patient treated | 1.661 | 1.775 | **1.449** | 0.806 | 0.295 | 0.121 |
| | **BOIN** | Probability selection | 0.071 | 0.201 | **0.28** | 0.233 | 0.134 | 0.073 |
| | | #of patient treated | 7.002 | 8.208 | **7.211** | 4.504 | 2.001 | 0.806 |

In scenario 1, the third dose was the MTD, and all designs selected the MTD with very similar probabilities. In particular, the CRM had the highest selection percentage of 31.8% for the MTD and mTPI had the lowest selection percentage of 26.3%. The selection probabilities for the MTD were very low for the designs.

Table 4.8 Comparison of scenario II with a toxicity target 25%

| | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | 0.01 | 0.02 | 0.03 | 0.04 | 0.1 | **0.25** |
| **Scenario2** | **3+3** | Probability selection | 0.004 | 0.011 | 0.014 | 0.091 | 0.367 | **0.51** |
| | | #of patient treated | 3.099 | 3.207 | 3.282 | 3.528 | 4.518 | **3.804** |
| | **CRM** | Probability selection | 0 | 0 | 0.003 | 0.012 | 0.229 | **0.756** |
| | | #of patient treated | 3.123 | 3.285 | 3.414 | 3.675 | 5.874 | **10.62** |
| | **BMA-CRM** | Probability selection | 0 | 0 | 0 | 0.001 | 0.078 | **0.921** |
| | | #of patient treated | 3.1 | 3.2 | 3.2 | 3.5 | 4.5 | **12.5** |
| | **BCRM** | Probability selection | 0 | 0 | 0.004 | 0.03 | 0.098 | **0.868** |
| | | #of patient treated | 3.1 | 3.2 | 3.45 | 3.987 | 4.952 | **11.24** |
| | **mTPI** | Probability selection | 0.001 | 0 | 0.003 | 0.02 | 0.245 | **0.731** |
| | | #of patient treated | 3.228 | 3.333 | 3.738 | 3.894 | 6.282 | **9.528** |
| | **mTPI-2** | Probability selection | 0.001 | 0 | 0.003 | 0.023 | 0.29 | **0.683** |
| | | #of patient treated | 3.282 | 3.483 | 3.738 | 4.527 | 6.813 | **8.157** |
| | **BOIN** | Probability selection | 0 | 0.002 | 0.004 | 0.027 | 0.278 | **0.692** |
| | | #of patient treated | 3.3 | 3.5 | 3.7 | 4.6 | 6.6 | **8.2** |

In the scenario 2, the sixth dose level was the MTD, and the MTD selection percentage using the BMA-CRM was the best design among the others. The selection percentage of the MTD for all the designs was very high in this scenario comparison to the other scenarios. The reason might be the true toxicity rates because it started with low toxicity rate then increased significantly at the last toxicity rate. The BCRM showed the second best performance with the selection probability of 86.8%. The comparison of the designs in the scenario III is given in Table 4.9.

Table 4.9 Comparison of scenario III with a toxicity target 25%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | **0.25** | 0.35 | 0.4 | 0.5 | 0.6 | 0.7 |
| **Scenario3** | **3+3** | Probability selection | **0.352** | 0.327 | 0.291 | 0.025 | 0.001 | 0 |
| | | #of patient treated | **5.106** | 2.853 | 1.131 | 0.339 | 0.036 | 0.003 |
| | **CRM** | Probability selection | **0.556** | 0.245 | 0.182 | 0.016 | 0 | 0 |
| | | #of patient treated | **6.815** | 7.14 | 1.938 | 0.342 | 0.036 | 0.003 |
| | **BMA-CRM** | Probability selection | **0.45** | 0.37 | 0.15 | 0.029 | 0 | 0 |
| | | #of patient treated | **14.2** | 6 | 2.3 | 0.5 | 0.1 | 0 |
| | **BCRM** | Probability selection | **0.552** | 0.302 | 0.116 | 0.025 | 0.004 | 0.001 |
| | | #of patient treated | **6.755** | 6.441 | 2.114 | 1.482 | 0.004 | 0.003 |
| | **mTPI** | Probability selection | **0.432** | 0.241 | 0.268 | 0.059 | 0 | 0 |
| | | #of patient treated | **4.473** | 2.582 | 0.718 | 0.128 | 0.011 | 0 |
| | **mTPI-2** | Probability selection | **0.479** | 0.351 | 0.116 | 0.051 | 0 | 0 |
| | | #of patient treated | **4.656** | 2.3 | 0.694 | 0.157 | 0.013 | 0 |
| | **BOIN** | Probability selection | **0.612** | 0.217 | 0.091 | 0.08 | 0 | 0 |
| | | #of patient treated | **7.002** | 8.208 | 7.211 | 4.504 | 2.001 | 0.806 |

The scenario 3 had the MTD at the first dose level. The BOIN, the CRM, and the BCRM were the best designs with the selection probability of 61.2%, 55.6% and 55.2%, respectively, in this scenario. The 3+3 design had the lowest selection percentage of 35.2%. In this scenario, we implemented a safety rule and all of the designs were able to terminate the trial early.

The comparison of the true MTD selection in the scenario IV is given in Table 4.10. Surprisingly the selection rate for the scenario 4 was very low for all designs. Although the fourth dose was the MTD, all designs selected wrong dose level for this scenario.

All methods had the wrong selection of the MTD when the true toxicity rate was changed very slightly. Wrong selection of the true MTD can have undesirable results when patients are treated at sensitive doses. Therefore, if the doses given to the treated patient are progressing at very low increments, the phase 1 designs used in such cases should be carefully selected.

Table 4.10 Comparison of scenario IV with a toxicity target 25%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | 0.19 | 0.21 | 0.23 | **0.25** | 0.27 | 0.29 |
| **Scenario4** | **3+3** | Probability selection | 0.339 | 0.291 | 0.240 | **0.079** | 0.046 | 0.005 |
| | | #of patient treated | 4.725 | 3.405 | 2.289 | **1.488** | 0.906 | 0.42 |
| | **CRM** | Probability selection | 0.235 | 0.262 | 0.267 | **0.135** | 0.064 | 0.037 |
| | | #of patient treated | 11.79 | 8.043 | 5.232 | **2.514** | 1.167 | 0.549 |
| | **BMA-CRM** | Probability selection | 0.28 | 0.17 | 0.197 | **0.132** | 0.101 | 0.12 |
| | | #of patient treated | 9.2 | 6.4 | 4.6 | **3** | 1.7 | 1.4 |
| | **BCRM** | Probability selection | 0.218 | 0.225 | 0.239 | **0.134** | 0.064 | 0.12 |
| | | #of patient treated | 11.24 | 8.001 | 5.121 | **2.501** | 1.154 | 0.5 |
| | **mTPI** | Probability selection | 0.244 | 0.389 | 0.165 | **0.113** | 0.058 | 0.031 |
| | | #of patient treated | 13.1 | 8.349 | 4.569 | **1.983** | 0.912 | 0.351 |
| | **mTPI-2** | Probability selection | 0.341 | 0.25 | 0.173 | **0.131** | 0.064 | 0.041 |
| | | #of patient treated | 13.05 | 8.043 | 4.482 | **2.238** | 1.005 | 444 |
| | **BOIN** | Probability selection | 0.264 | 0.25 | 0.231 | **0.124** | 0.089 | 0.041 |
| | | #of patient treated | 12.81 | 7.601 | 4.9 | **2.5** | 1.2 | 0.5 |

The simulation result of the scenario V is given in Table 4.11. In the scenario 5, the selection percentage of the MTD for all designs was quite close, but the CRM had the highest selection rate with 45.5% and the BMA-CRM design has the second best performance. Similarly, the 3+3 design has the worst performance. However, the probability of selecting the previous dose as the MTD was very high in all other designs except the BMA-CRM. In the last scenario, the BMA-CRM was very robust. It produced the best MTD selection percentage among the other designs.

Table 4.11 Comparison of scenario V with a toxicity target 25%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | 0.01 | 0.1 | 0.2 | **0.25** | 0.3 | 0.35 |
| **Scenario5** | **3+3** | Probability selection | 0.093 | 0.184 | 0.242 | **0.297** | 0.109 | 0.074 |
| | | #of patient treated | 3.336 | 4.524 | 4.365 | **3.069** | 2.01 | 0.813 |
| | **CRM** | Probability selection | 0.001 | 0.066 | 0.207 | **0.455** | 0.195 | 0.076 |
| | | #of patient treated | 3.39 | 5.934 | 9.393 | **10.96** | 3.345 | 1.338 |
| | **BMA-CRM** | Probability selection | 0 | 0.06 | 0.11 | **0.4** | 0.2 | 0.23 |
| | | #of patient treated | 8.1 | 7.3 | 5.9 | **10.8** | 2 | 0.9 |
| | **BCRM** | Probability selection | 0.003 | 0.076 | 0.247 | **0.374** | 0.154 | 0.146 |
| | | #of patient treated | 3.356 | 6.1 | 8.541 | **9.604** | 2.987 | 1.524 |
| | **mTPI** | Probability selection | 0.005 | 0.105 | 0.283 | **0.387** | 0.148 | 0.072 |
| | | #of patient treated | 3.423 | 7.305 | 10.37 | **10.52** | 2.316 | 1.008 |
| | **mTPI-2** | Probability selection | 0.005 | 0.034 | 0.37 | **0.363** | 0.166 | 0.062 |
| | | #of patient treated | 4.458 | 8.343 | 8.796 | **9.815** | 2.301 | 0.987 |
| | **BOIN** | Probability selection | 0.002 | 0.041 | 0.104 | **0.362** | 0.253 | 0.208 |
| | | #of patient treated | 3.3 | 5.6 | 7.7 | **9.75** | 4.3 | 2.4 |

Table 4.12 Comparison of scenario VI with a toxicity target 25%

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | 0.02 | 0.04 | 0.08 | 0.12 | 0.17 | **0.25** |
| **Scenario6** | **3+3** | Probability selection | 0.019 | 0.071 | 0.134 | 0.193 | 0.232 | **0.351** |
| | | #of patient treated | 3.183 | 3.525 | 3.915 | 3.975 | 3.669 | **4.472** |
| | **CRM** | Probability selection | 0 | 0.003 | 0.031 | 0.150 | 0.341 | **0.477** |
| | | #of patient treated | 3.261 | 3.717 | 4.668 | 5.649 | 6.51 | **6.195** |
| | **BMA-CRM** | Probability selection | 0 | 0 | 0.01 | 0.01 | 0.18 | **0.766** |
| | | #of patient treated | 3.3 | 3.5 | 4 | 4.7 | 5.3 | **9.7** |
| | **BCRM** | Probability selection | 0 | 0.003 | 0.004 | 0.149 | 0.161 | **0.677** |
| | | #of patient treated | 3.275 | 3.705 | 4.117 | 5.202 | 5.941 | **6.874** |
| | **mTPI** | Probability selection | 0.001 | 0.004 | 0.051 | 0.208 | 0.281 | **0.455** |
| | | #of patient treated | 3.465 | 4.864 | 5.628 | 6.231 | 5.598 | **5.214** |
| | **mTPI-2** | Probability selection | 0.001 | 0.002 | 0.058 | 0.185 | 0.337 | **0.417** |
| | | #of patient treated | 3.588 | 4.386 | 5.733 | 6.129 | 5.391 | **4.773** |
| | **BOIN** | Probability selection | 0.001 | 0.07 | 0.055 | 0.108 | 0.364 | **0.406** |
| | | #of patient treated | 3.6 | 4.5 | 5.7 | 6.1 | 5.4 | **4.8** |

Now the true selection probabilities of the MTD for the scenario I and scenario II can be seen in Figure 4.4.

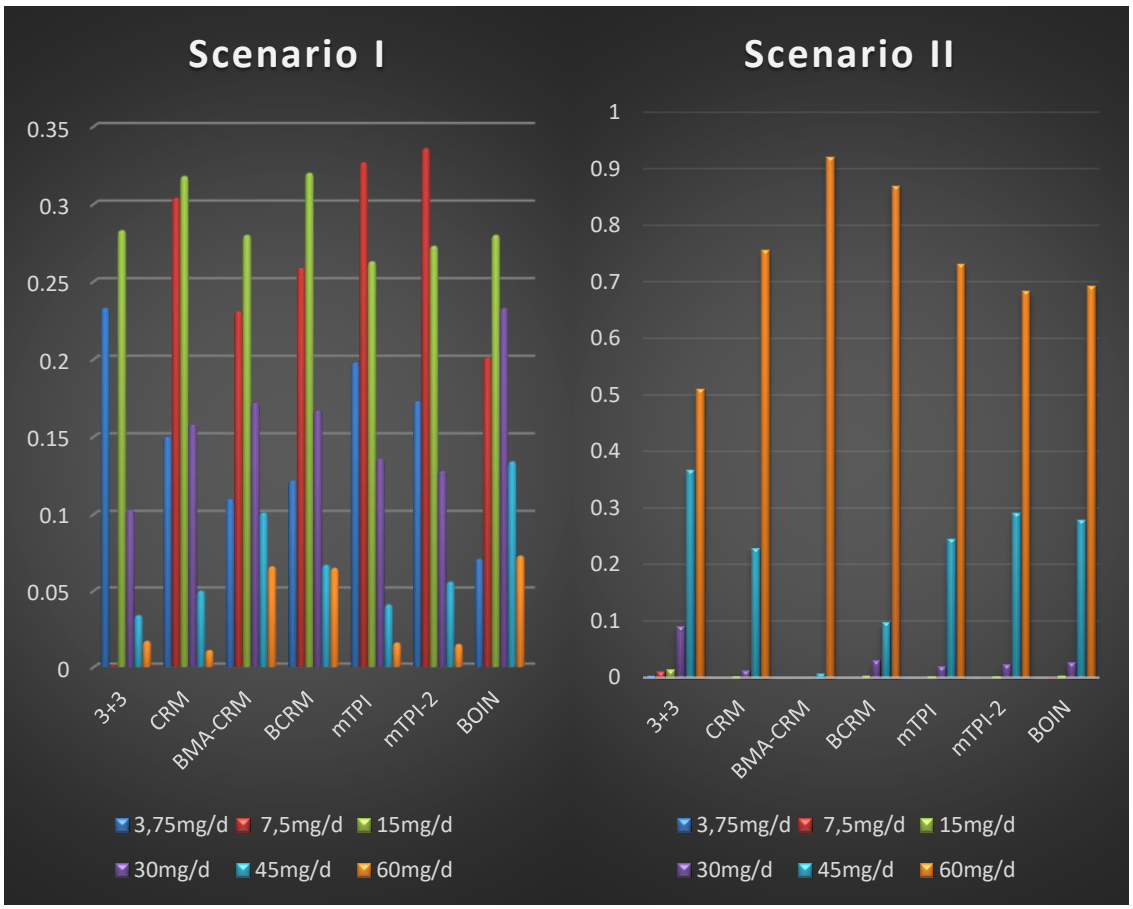


Figure 4.12 Comparison of the designs for scenario I and scenario II (Story 2)

It is clear from the histogram of the first scenario that selection probability of the second dose and the third dose were very close in all designs except the 3+3. However, the true MTD dose was the third dose.

The BOIN, CRM, BMA-CRM, and BCRM have selected the correct dose level as the MTD. In contrast, the mTPI and mTPI -2 were selected the wrong dose level as the MTD. This result was unexpected because, in almost all scenarios, the mTPI and mTPI -2 were selected the true MTD dose level.

It is widely accepted that the phase I experiments are small. However, the results of this methodology have been poorly addressed in the literature. A small phase I trials usually provide the lowest recommended dose for the phase II, which results in low efficacy or high toxicity when the recommended doses are too low or too high. For the phase I studies, appropriate sample sizes need to be discussed and investigated further. The reason behind this wrong selection might be the small sample size.

Figure 4.2 shows the MTD selection of the seven designs for the scenario III and scenario IV. Remarkably, the selection probabilities of the MTD for the first dose in scenario III were high in all doses. For the scenario IV, all designs selected wrong dose level as the MTD. The reason might be because of the true toxicity rates were changed very slightly.

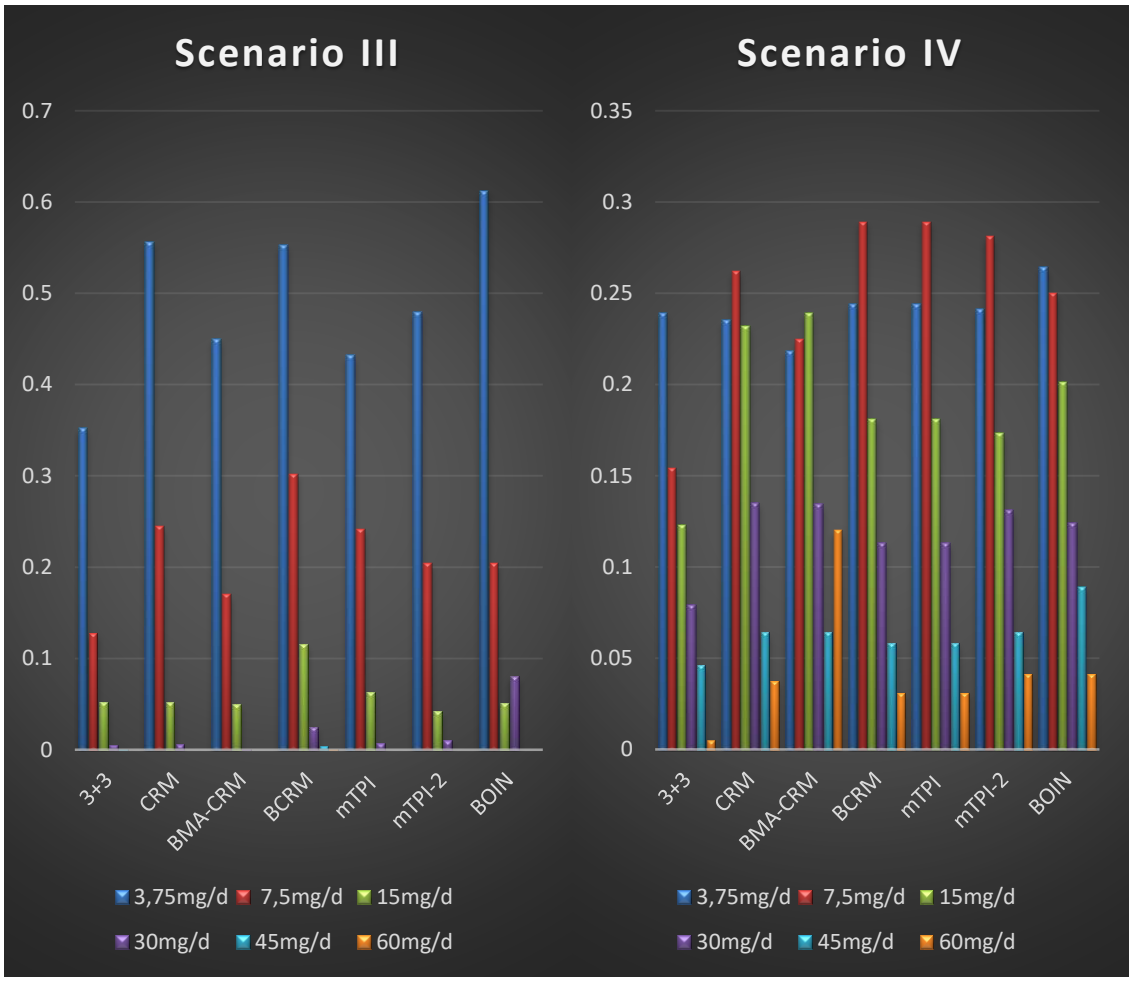


Figure 4.13 Comparison of the designs for scenario III and scenario IV (Story 2)

The MTD selection of the scenario V and VI can be seen in Figure 4.6. Although the fourth dose was the MTD, the mTPI -2 was selected the third dose level as the MTD in the scenario V. In the scenario VI, all designs were selected the correct dose level as the MTD.

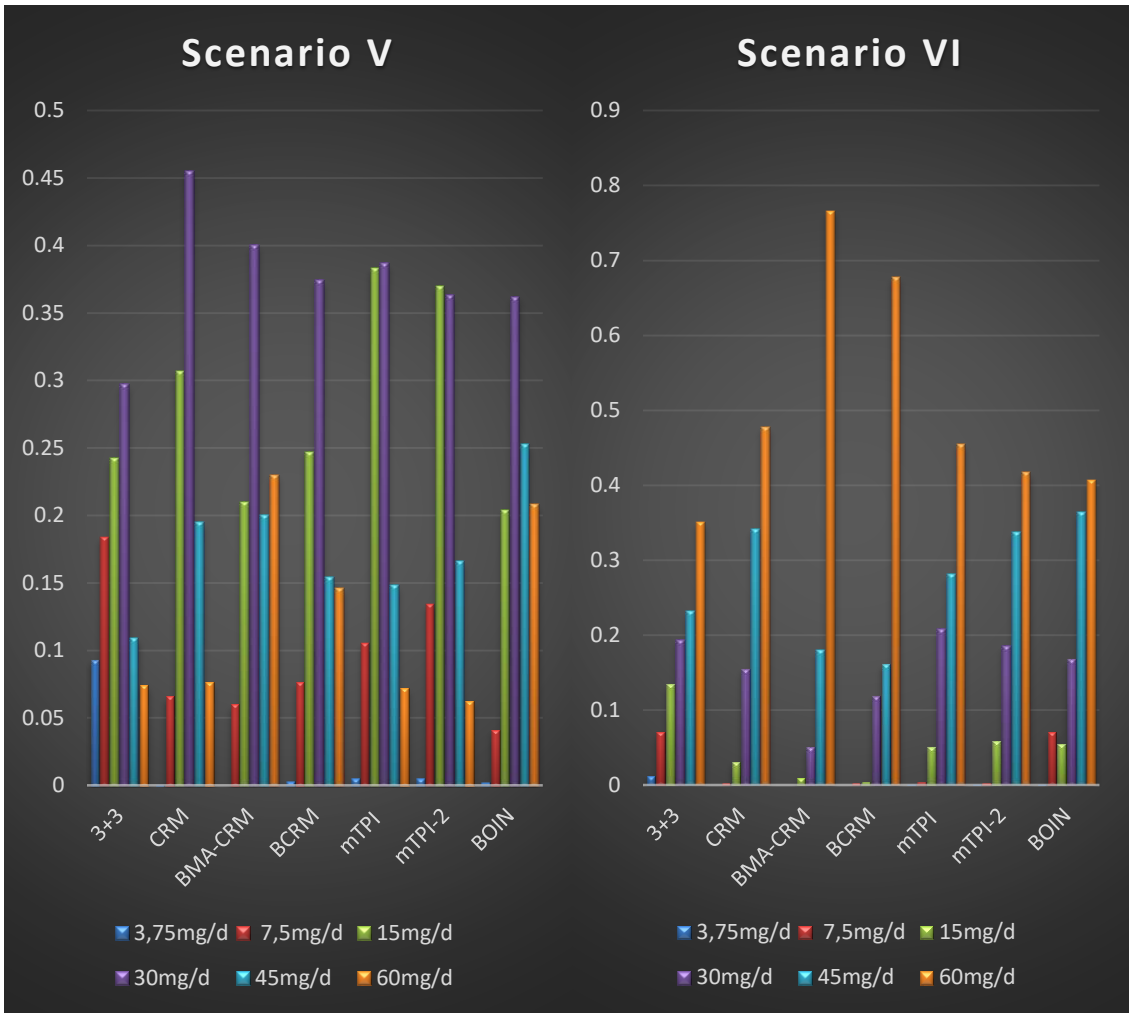


Figure 4.14 Comparison of the designs for scenario V and scenario VI (Story 2)

## 4.3 Simulation Study for Scenario (Story) III

In the last simulation study, the designs that showed the best performances in the previous simulation studies and the designs obtained by using different model structures and preliminaries in the CRM method were compared. The comparisons were made with eight different scenarios and performances in different cases were observed in each simulation run. In the model-based methods, skeleton choices have been made carefully because of the significant effect of the skeletons on correct MTD selection.

In some scenarios, the results were observed when the initial guesses were not close to the true toxicity rate. It has been investigated in which conditions the designs using these eight different scenarios produce better results.

For the CRM model, all skeletons were assigned one by one to the models and the best-performed one was used for comparison. The first skeleton is for the case where the toxicity starts from a certain level and progresses gradually to medium levels. In the

second skeleton, the toxicity started at a very low level then rapidly increased to toxic levels. Similarly, the third skeleton was started at a specific dose and fixed at the doses around the MTD.

The fourth skeleton, unlike the other skeletons, quickly rose to toxic doses after starting at the MTD level. The last dose level in the last skeleton increased to the MTD shortly after the dose level started at a low level, and then, progressed rapidly to toxic levels. Table 4.13-4.20 shows the simulation results of the scenarios for the CRM design where different model structures and priors used, BMA-CRM, mTPI and mTPI-2. The probability selection of the MTD included in the tables. We carried out 10,000 simulations for each scenario.

Table 4.13 Comparison of scenario I with a toxicity target 30% (Story 3)

| | | | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.13 | 0.18 | 0.22 | 0.3 | 0.36 | 0.44 |
| **Scenario1** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | 0.151 | 0.185 | 0.188 | **0.212** | 0.159 | 0.105 |
| | | | **Uniform** | Probability selection | 0.139 | 0.144 | 0.192 | **0.211** | 0.181 | 0.133 |
| | | | **Lognormal** | Probability selection | 0.167 | 0.181 | 0.211 | **0.215** | 0.134 | 0.091 |
| | | **Logistic** | **Gamma** | Probability selection | 0.167 | 0.182 | 0.194 | **0.214** | 0.136 | 0.108 |
| | | | **Uniform** | Probability selection | 0.195 | 0.162 | 0.188 | **0.198** | 0.136 | 0.122 |
| | | | **Lognormal** | Probability selection | 0.217 | 0.175 | 0.202 | **0.177** | 0.138 | 0.091 |
| | | **Power** | **Gamma** | Probability selection | 0.146 | 0.182 | 0.198 | **0.216** | 0.158 | 0.099 |
| | | | **Uniform** | Probability selection | 0.138 | 0.144 | 0.189 | **0.213** | 0.178 | 0.138 |
| | | | **Lognormal** | Probability selection | 0.195 | 0.17 | 0.204 | **0.206** | 0.138 | 0.087 |
| | **BMA-CRM** | | | Probability selection | 0.032 | 0.171 | 0.283 | **0.288** | 0.15 | 0.07 |
| | **mTPI-2** | | | Probability selection | 0.065 | 0.162 | 0.253 | **0.256** | 0.201 | 0.059 |
| | **mTPI** | | | Probability selection | 0.043 | 0.129 | 0.225 | **0.287** | 0.255 | 0.057 |

In the first scenario, all designs were selected the MTD correctly. It should be noted here that the selection probabilities for all designs are very low. The reason for this is

that the true toxicity probabilities in the scenario are very close to the MTD. Despite this, all designs have selected the MTD correctly.

Another noteworthy situation is that there are differences in selection probabilities when the different model structures and different priors in the CRM are considered. However, this does not affect the selection of the correct MTD. Among these differences, the best selection probability was obtained when the power model and gamma prior were chosen in the CRM. The best selection probability produced among designs was the BMA-CRM method with 28.8%. However, the selection probability of the previous dose level was very close to the MTD selection probability.

Table 4.14 Comparison of scenario II with a toxicity target 30% (Story 3)

| | | | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.01 | 0.05 | 0.3 | 0.65 | 0.72 | 0.85 |
| **Scenario2** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | 0.1 | 0.16 | **0.653** | 0.086 | 0 | 0 |
| | | | **Uniform** | Probability selection | 0.1 | 0.126 | **0.746** | 0.027 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.103 | 0.165 | **0.661** | 0.07 | 0.001 | 0 |
| | | **Logistic** | **Gamma** | Probability selection | 0.103 | 0.161 | **0.629** | 0.106 | 0.001 | 0 |
| | | | **Uniform** | Probability selection | 0.104 | 0.154 | **0.616** | 0.125 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.103 | 0.173 | **0.617** | 0.106 | 0.001 | 0 |
| | | **Power** | **Gamma** | Probability selection | 0.1 | 0.151 | **0.662** | 0.086 | 0 | 0 |
| | | | **Uniform** | Probability selection | 0.1 | 0.126 | **0.744** | 0.029 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.102 | 0.162 | **0.654** | 0.08 | 0.001 | 0 |
| | **BMA-CRM** | | | Probability selection | 0 | 0.051 | **0.91** | 0.04 | 0 | 0 |
| | **mTPI-2** | | | Probability selection | 0.001 | 0.306 | **0.685** | 0.007 | 0.001 | 0 |
| | **mTPI** | | | Probability selection | 0 | 0.334 | **0.66** | 0.005 | 0.001 | 0 |

In the scenario 2, the BMA-CRM is the design that clearly produced the best performance on selecting the MTD correctly. The BMA-CRM produces very good results if the difference between the correct MTD dose and the previous-subsequent

dose is greater. Although the other designs produced similar results, the best choice was obtained when the hyperbolic tangent model and a uniform prior were selected.

This scenario demonstrates how different model structures and priors are effective in selecting the right MTD. The selection probability of choosing the right MTD for the BMA-CRM is as high as 91%. The second best performance was obtained with 74.6% when the hyperbolic tangent model and a uniform prior were selected in the CRM.

Table 4.15 Comparison of scenario III with a toxicity target 30% (Story 3)

| | | | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.3 | 0.55 | 0.65 | 0.76 | 0.84 | 0.92 |
| **Scenario3** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | **0.849** | 0.145 | 0.006 | 0 | 0 | 0 |
| | | | **Uniform** | Probability selection | **0.799** | 0.201 | 0 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | **0.843** | 0.15 | 0.006 | 0 | 0 | 0 |
| | | **Logistic** | **Gamma** | Probability selection | **0.805** | 0.183 | 0.01 | 0 | 0 | 0 |
| | | | **Uniform** | Probability selection | **0.783** | 0.204 | 0.012 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | **0.829** | 0.159 | 0.01 | 0 | 0 | 0 |
| | | **Power** | **Gamma** | Probability selection | **0.826** | 0.167 | 0.007 | 0 | 0 | 0 |
| | | | **Uniform** | Probability selection | **0.797** | 0.203 | 0 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | **0.837** | 0.155 | 0.008 | 0 | 0 | 0 |
| | **BMA-CRM** | | | Probability selection | **0.57** | 0.08 | 0 | 0 | 0 | 0 |
| | **mTPI-2** | | | Probability selection | **0.706** | 0.265 | 0.002 | 0 | 0 | 0 |
| | **mTPI** | | | Probability selection | **0.663** | 0.332 | 0 | 0 | 0 | 0 |

In the scenario 3, where the correct MTD was in the first dose, the selection probability of the BMA-CRM was very low compared to the other designs, although all doses select the correct MTD. Other designs have selected the correct MTD with high selection probabilities.

The best selection probability was produced by the CRM method when the model structure and prior were the hyperbolic tangent and gamma respectively. The highest selection probability of the correct MTD selection was obtained with 84.9% when the hyperbolic tangent model and gamma prior were selected in the CRM. In the case in which the correct MTD was the first dose, the BMA-CRM has the worst performance. In this scenario, the importance of the model structures and priors was seen when selection probabilities were examined.

Table 4.16 Comparison of scenario IV with a toxicity target 30% (Story 3)

| | | | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.04 | 0.3 | 0.68 | 0.75 | 0.89 | 0.97 |
| **Scenario4** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | 0.149 | **0.77** | 0.08 | 0 | 0 | 0 |
| | | | **Uniform** | Probability selection | 0.137 | **0.834** | 0.029 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.168 | **0.762** | 0.069 | 0 | 0 | 0 |
| | | **Logistic** | **Gamma** | Probability selection | 0.181 | **0.724** | 0.093 | 0 | 0 | 0 |
| | | | **Uniform** | Probability selection | 0.182 | **0.728** | 0.087 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.183 | **0.737** | 0.077 | 0.002 | 0 | 0 |
| | | **Power** | **Gamma** | Probability selection | 0.15 | **0.769** | 0.08 | 0 | 0 | 0 |
| | | | **Uniform** | Probability selection | 0.136 | **0.834** | 0.03 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.169 | **0.753** | 0.077 | 0 | 0 | 0 |
| | **BMA-CRM** | | | Probability selection | 0.052 | **0.94** | 0.02 | 0 | 0 | 0 |
| | **mTPI-2** | | | Probability selection | 0.293 | **0.703** | 0.002 | 0.001 | 0 | 0 |
| | **mTPI** | | | Probability selection | 0.343 | **0.655** | 0.002 | 0 | 0 | 0 |

In the scenario 4, the 2nd dose was determined as the MTD. This dose has chosen by all designs with high selection probabilities. The most important and similar result in this scenario was that when the difference between the right MTD and the previous-subsequent dose was high, the BMA-CRM was selected the correct MTD with a high selection probability.

In such scenarios, the BMA-CRM was once again seen as the most effective design. Among other designs, the second most accurate MTD selection probability was generated in the CRM method when the model structure was hyperbolic tangent and prior was uniform or the model structure was power and prior was uniform. In all designs, the selection probability of the next dose as the MTD was almost 0%.

Table 4.17 Comparison of scenario V with a toxicity target 30% (Story 3)

| | | | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.01 | 0.05 | 0.1 | 0.6 | 0.7 | 0.9 |
| **Scenario5** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | 0.1 | 0.134 | 0.472 | 0.294 | 0 | 0 |
| | | | **Uniform** | Probability selection | 0.1 | 0.106 | 0.589 | 0.204 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.103 | 0.132 | 0.509 | 0.251 | 0 | 0 |
| | | **Logistic** | **Gamma** | Probability selection | 0.103 | 0.13 | 0.469 | 0.292 | 0.006 | 0 |
| | | | **Uniform** | Probability selection | 0.102 | 0.126 | 0.439 | 0.333 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.104 | 0.134 | 0.468 | 0.29 | 0.005 | 0 |
| | | **Power** | **Gamma** | Probability selection | 0.1 | 0.134 | 0.501 | 0.265 | 0 | 0 |
| | | | **Uniform** | Probability selection | 0.1 | 0.107 | 0.601 | 0.192 | 0 | 0 |
| | | | **Lognormal** | Probability selection | 0.104 | 0.136 | 0.488 | 0.266 | 0 | 0 |
| | **BMA-CRM** | | | Probability selection | 0 | 0 | 0.65 | 0.35 | 0 | 0 |
| | **mTPI-2** | | | Probability selection | 0.001 | 0.009 | 0.958 | 0.03 | 0.002 | 0 |
| | **mTPI** | | | Probability selection | 0 | 0.006 | 0.969 | 0.025 | 0 | 0 |

In the fifth scenario, none of the true toxicity probabilities is 0.3, so it is expected that the dose to be selected as the correct MTD will be selected as the dose closest to the toxicity target of 0.3. The closest dose to the true toxicity probability is the third dose. Designs that produced the highest selection probability in the third dose selection probability were 96.9% and 95.8%, respectively, in mTPI and mTPI-2.

In all the other designs, even the third dose is selected as MTD, selection probability of the next dose level was high. The probability of the next dose is 0.6, which means that the patient is treated at the dose level where the toxicity was high. The most important result in this scenario was that, in the absence of the toxicity probability, mTPI and mTPI-2 designs produced more accurate and effective results because they made an estimation based on intervals.

Table 4.18 Comparison of scenario VI with a toxicity target 30% (Story 3)

| | | | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.25 | 0.27 | 0.29 | 0.31 | 0.33 | 0.35 |
| **Scenario6** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | 0.303 | 0.236 | 0.167 | 0.141 | 0.088 | 0.065 |
| | | | **Uniform** | Probability selection | 0.255 | 0.173 | 0.195 | 0.168 | 0.122 | 0.087 |
| | | | **Lognormal** | Probability selection | 0.43 | 0.167 | 0.167 | 0.108 | 0.067 | 0 |
| | | **Logistic** | **Gamma** | Probability selection | 0.433 | 0.178 | 0.138 | 0.113 | 0.075 | 0 |
| | | | **Uniform** | Probability selection | 0.405 | 0.179 | 0.142 | 0.119 | 0.079 | 0 |
| | | | **Lognormal** | Probability selection | 0.445 | 0.174 | 0.156 | 0.097 | 0.067 | 0 |
| | | **Power** | **Gamma** | Probability selection | 0.287 | 0.22 | 0.166 | 0.153 | 0.096 | 0 |
| | | | **Uniform** | Probability selection | 0.266 | 0.172 | 0.195 | 0.159 | 0.114 | 0.094 |
| | | | **Lognormal** | Probability selection | 0.418 | 0.186 | 0.169 | 0.11 | 0.064 | 0.052 |
| | **BMA-CRM** | | | Probability selection | 0.18 | 0.33 | 0.16 | 0.13 | 0.08 | 0.08 |
| | **mTPI-2** | | | Probability selection | 0.271 | 0.177 | 0.148 | 0.145 | 0.172 | 0.082 |
| | **mTPI** | | | Probability selection | 0.208 | 0.18 | 0.151 | 0.192 | 0.157 | 0.11 |

In the scenario 6, where there was no target toxicity as 0.3, the doses were very slightly increased after starting at a certain level. In such scenarios, it is quite difficult to determine the correct MTD. None of the designs chose the right MTD. The results obtained were not suitable for the MTD selection.

In the clinical trials, where the sensitive doses need to be used, the dose response methods may not give the correct result. In these cases, the trial to be conducted may

need to be well-planned. In the scenario, all designs except the BMA-CRM were mistakenly chosen as the first dose level as the MTD. In addition, the BMA-CRM incorrectly selected the second dose level as the MTD. In such scenarios, the selection probabilities produced by these designs may not be reliable.

Table 4.19 Comparison of scenario VII with a toxicity target 30% (Story 3)

| | | | | | **Dose Levels** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 | 0.55 |
| **Scenario7** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | **0.699** | 0.242 | 0.056 | 0.002 | 0 | 0 |
| | | | **Uniform** | Probability selection | **0.636** | 0.364 | 0 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | **0.697** | 0.227 | 0.065 | 0.009 | 0 | 0 |
| | | **Logistic** | **Gamma** | Probability selection | **0.655** | 0.253 | 0.076 | 0.014 | 0.001 | 0 |
| | | | **Uniform** | Probability selection | **0.599** | 0.283 | 0.118 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | **0.684** | 0.231 | 0.067 | 0.016 | 0.001 | 0 |
| | | **Power** | **Gamma** | Probability selection | **0.67** | 0.27 | 0.057 | 0.003 | 0 | 0 |
| | | | **Uniform** | Probability selection | **0.656** | 0.344 | 0 | 0 | 0 | 0 |
| | | | **Lognormal** | Probability selection | **0.666** | 0.252 | 0.068 | 0.012 | 0 | 0 |
| | **BMA-CRM** | | | Probability selection | **0.581** | 0.252 | 0.143 | 0.014 | 0.013 | 0 |
| | **mTPI-2** | | | Probability selection | **0.629** | 0.195 | 0.1 | 0.05 | 0.022 | 0.002 |
| | **mTPI** | | | Probability selection | **0.516** | 0.251 | 0.126 | 0.07 | 0.025 | 0.002 |

In the scenario 7, where the first dose was determined as the MTD and the subsequent doses gradually increased at short intervals, the best-performed method was the CRM design, where the model structure is the hyperbolic tangent and the prior was gamma. All designs chose the correct dose level as the MTD and the selection probabilities were similar.

In this scenario, there is no significant difference between the designs. However, it appears clearly in this scenario that the selection of the wrong model structure and the priors can change the selection probabilities.

Table 4.20 Comparison of scenario VIII with a toxicity target 30% (Story 3)

| | | | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **1mg** | **2 mg** | **4 mg** | **8 mg** | **12 mg** | **16 mg** |
| | **Method** | **MS** | **PD** | True toxicity rate | 0.15 | 0.18 | 0.21 | 0.24 | 0.27 | 0.3 |
| **Scenario8** | **CRM** | **Hyperbolic-tangent** | **Gamma** | Probability selection | 0.153 | 0.166 | 0.15 | 0.184 | 0.153 | **0.194** |
| | | | **Uniform** | Probability selection | 0.129 | 0.125 | 0.156 | 0.175 | 0.173 | **0.242** |
| | | | **Lognormal** | Probability selection | 0.129 | 0.152 | 0.176 | 0.153 | 0.172 | **0.217** |
| | | **Logistic** | **Gamma** | Probability selection | 0.15 | 0.161 | 0.143 | 0.152 | 0.174 | **0.22** |
| | | | **Uniform** | Probability selection | 0.151 | 0.152 | 0.166 | 0.139 | 0.194 | **0.198** |
| | | | **Lognormal** | Probability selection | 0.15 | 0.15 | 0.178 | 0.137 | 0.174 | **0.212** |
| | | **Power** | **Gamma** | Probability selection | 0.145 | 0.166 | 0.151 | 0.176 | 0.166 | **0.196** |
| | | | **Uniform** | Probability selection | 0.133 | 0.126 | 0.152 | 0.173 | 0.181 | **0.236** |
| | | | **Lognormal** | Probability selection | 0.145 | 0.172 | 0.162 | 0.136 | 0.163 | **0.222** |
| | **BMA-CRM** | | | Probability selection | 0.04 | 0.152 | 0.211 | 0.151 | 0.162 | **0.261** |
| | **mTPI-2** | | | Probability selection | 0.071 | 0.129 | 0.165 | 0.173 | 0.183 | **0.267** |
| | **mTPI** | | | Probability selection | 0.049 | 0.101 | 0.12 | 0.163 | 0.243 | **0.312** |

In the scenario 8, where the target toxicity probability is at the final dose level, the dose levels gradually increase from the start of a certain level until the target toxicity was reached. The reason for the low selection probabilities may be that the dose levels were close to the target toxicity. In this scenario, the design with the best performance was the mTPI with 31.2%. All designs chose the MTD correctly.

The CRM design with the best model structure and prior distribution was hyperbolic tangent model and uniform prior distribution. It may be difficult to examine the comparatives in each scenario from the tables. Hence, the histogram graphs for the scenarios are given in Figures 4.7-4.17. With these graphs, the results of the phase 1 dose-response methods can be easily seen.

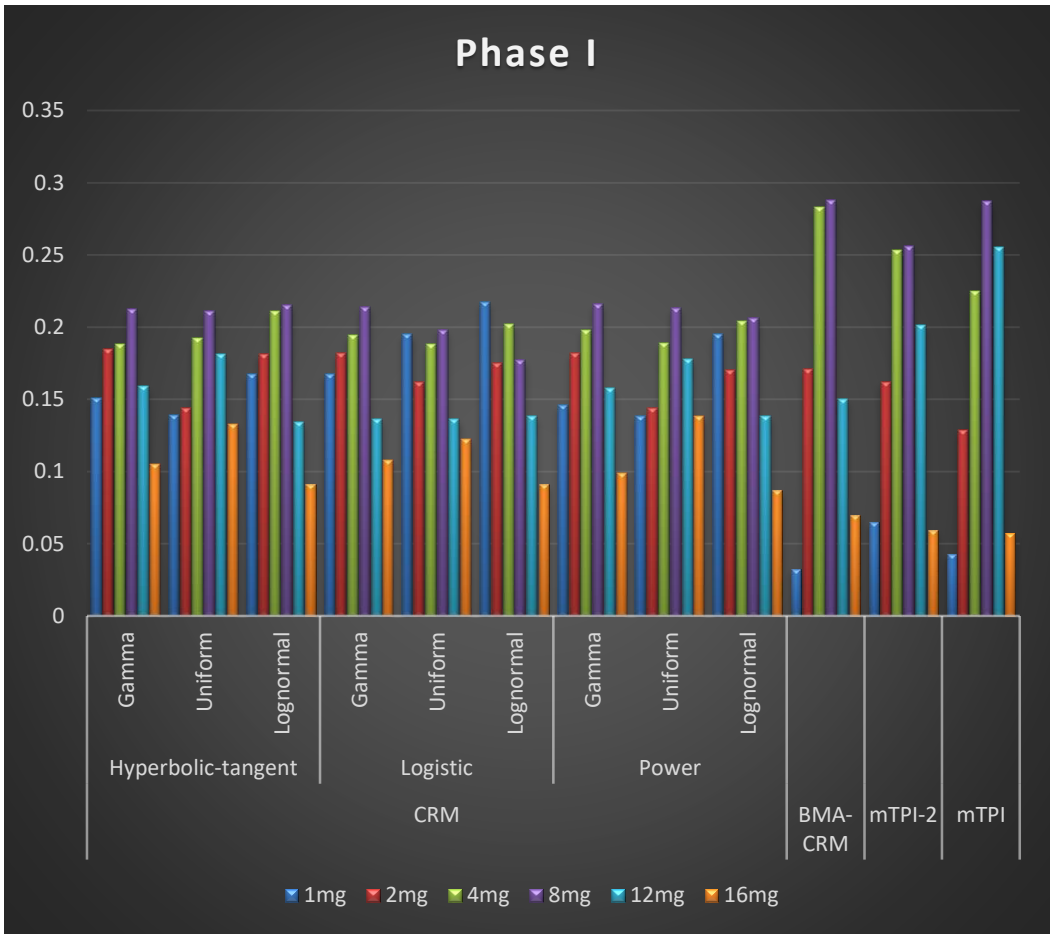


Figure 4.15 Comparison of the designs for scenario I (Story 3)

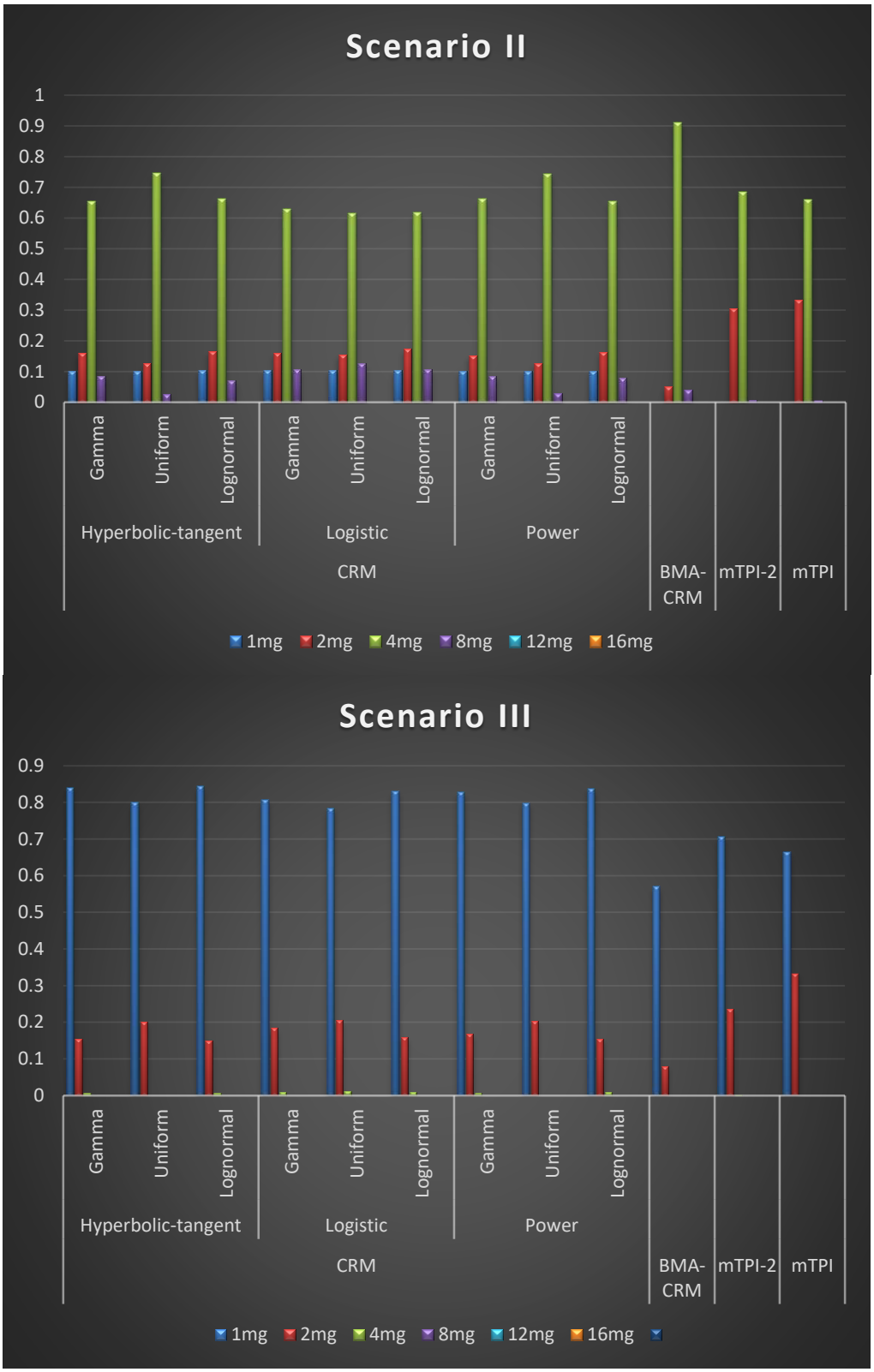


Figure 4.16 Comparison of the designs for scenario II- III (Story 3)

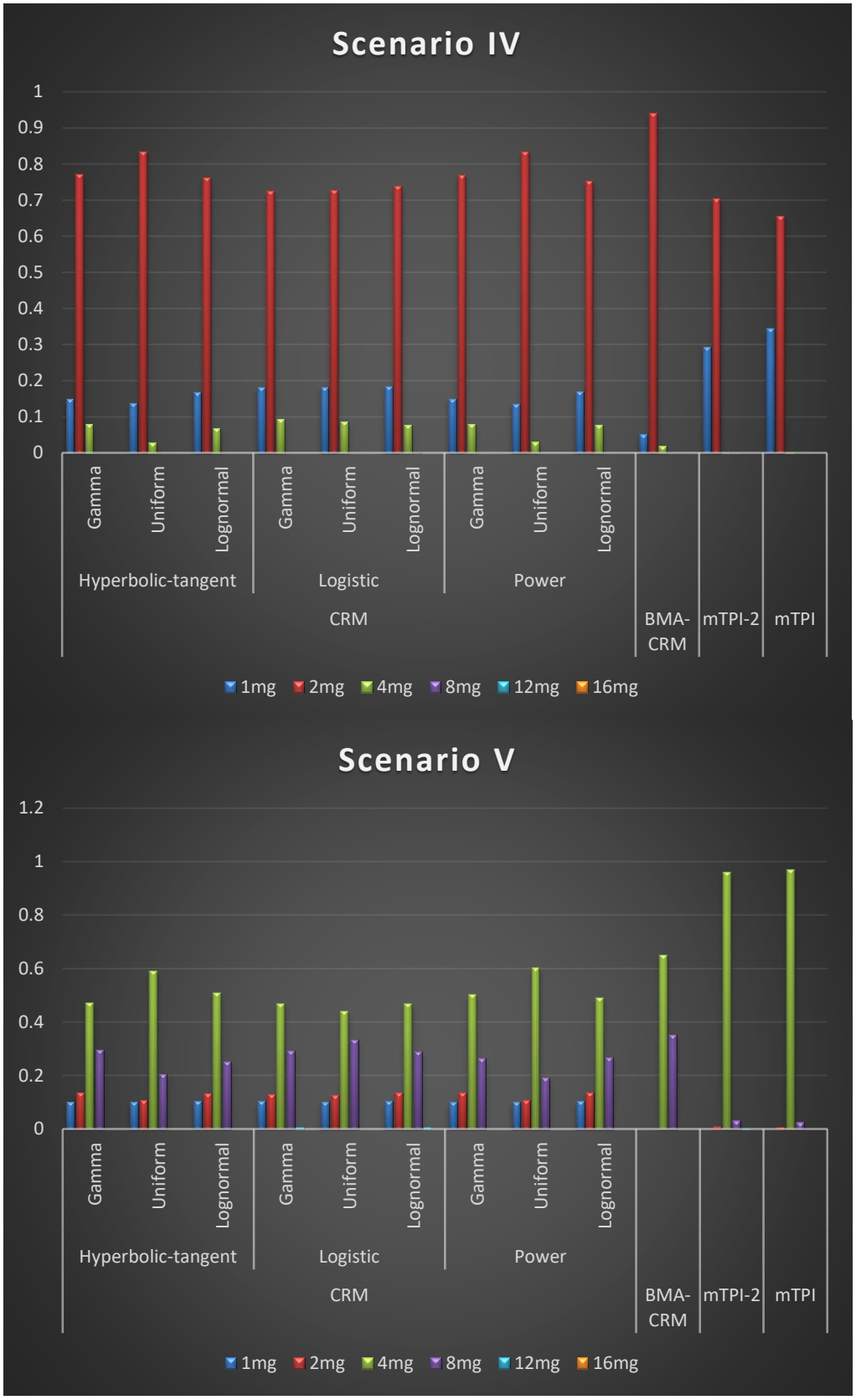


Figure 4.17 Comparison of the designs for scenario IV- V (Story 3)

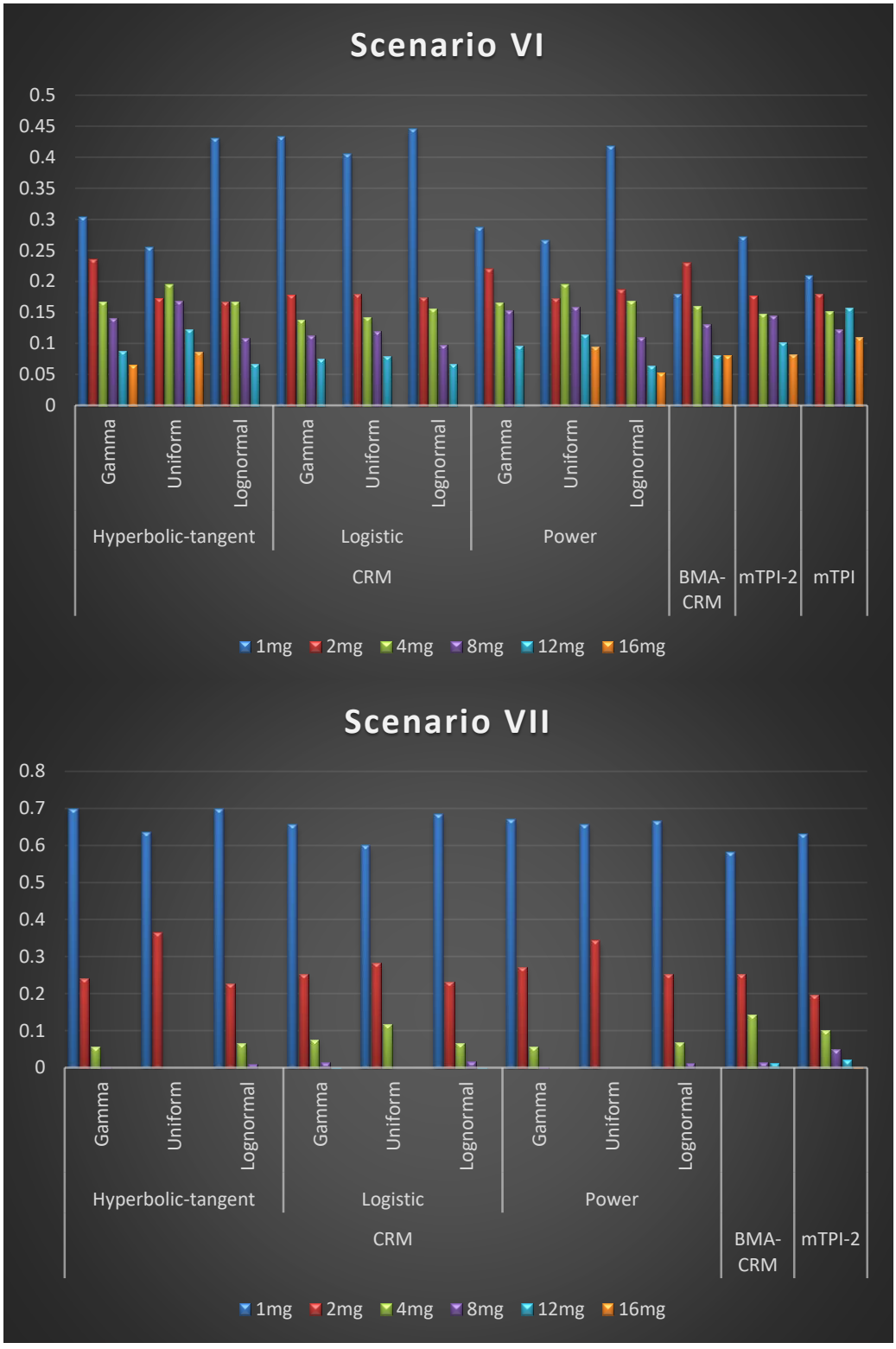


Figure 4.18 Comparison of the designs for scenario VI- VII (Story 3)

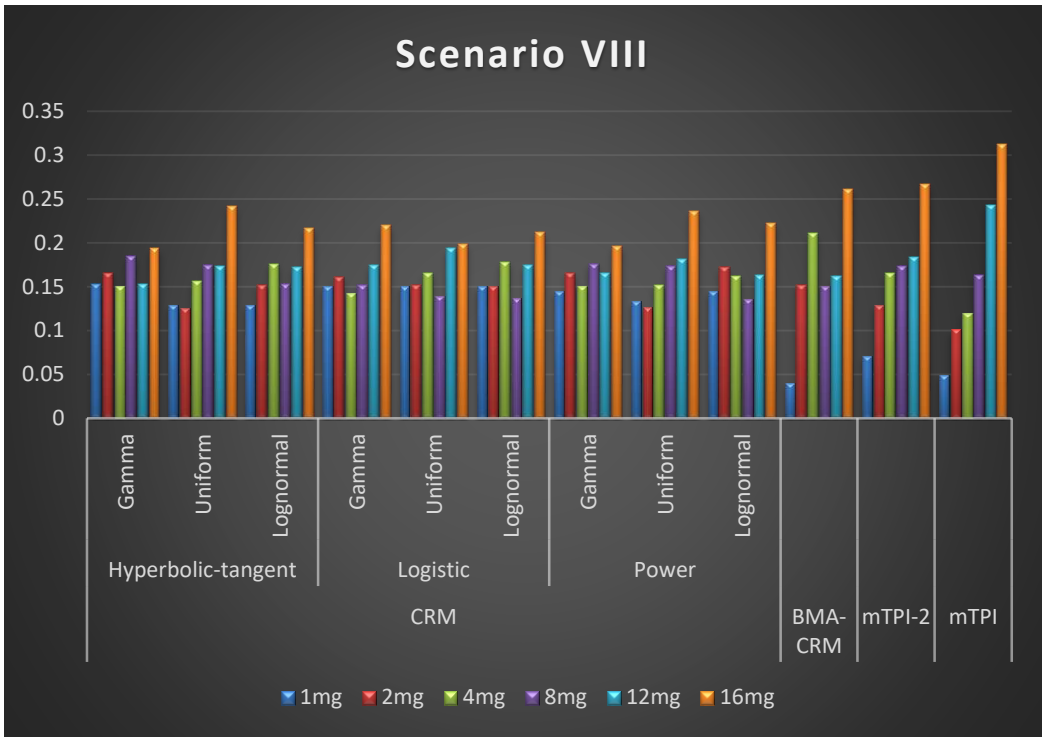


Figure 4.19 Comparison of the designs for scenario VIII (Story 3)

# CHAPTER 5

# CONCLUSION

This chapter summarizes the methods and the simulation results that were presented in this thesis. Each chapter will be described in detail in the next sections.

## 5.1 Summary of the Chapters

### 5.1.1 Chapter 1

Chapter 1 discusses the importance and effects of the clinical trials in the drug development process. The purpose of drug development is to help the patients overcome their illnesses and improve their quality of lives. The drug development process is designed to ensure that innovative medicines are effective, safe and have treatments that can be brought to the patient's use as soon as possible. Chapter 1 discusses the factors needed to make these designs more effective. In addition, the stages of the four different phases of the clinical trials are summarized.

In the next section of the Chapter 1, it is mentioned that what the terms mean in the clinical trials, and at what stages of the clinical trials these terms are used. This section clearly describes how a protocol to be created, how to determine the sample size of the attempted experiment, how data should be collected, what methods are applied at this stage, and what pharmacokinetic and pharmacodynamics terms mean. In addition, the purpose of the thesis is summarized in this chapter.

### 5.1.2 Chapter 2

In this chapter, seven different design algorithms and theories used in the thesis are explained. In recent years, the number of traditional methods and the dose response

methods in the phase I clinical trials have become popular. These methods are very necessary in terms of cost effectiveness and time-saving. However, the physical applicability of implementing these effective methods to determine the MTD has been still discussed. The theory of the most used methods in the literature and their differences are explained clearly in this chapter. The methods used are theoretically compared and the advantages and disadvantages are pointed out.

### 5.1.3 Chapter 3

In Chapter 3, the real life stories for each simulation run are presented. These stories have been adapted from the phase I clinical trial projects. For each story, the number and probabilities of the skeletons, the dose levels, the name of the drug, the number of patients included in the trial, the number of the simulation runs, the age range of the patients, the sexes of the patients and inhibitors of the drug have been given.

### 5.1.4 Chapter 4

In order to investigate the performances of the phase I dose-response methods, the chapter 4 implemented a simulation study which examines the phase I methods that have been used in this dissertation. Three different real life stories and a total of 20 simulation scenarios were used in this thesis. In the thesis's first simulation study, the Story 1 and the Story 2 were used to comparing the most commonly used phase-1 clinical trial dose response methods. In the first part, the prominent methods were the CRM, BMA-CRM, mTPI and mTPI -2. The traditional 3 + 3 design was the worst performing method in the scenarios that we have created.

In the second simulation study of the thesis, the story 3 compared the prominent methods that appeared in the first section and the results of using different model structures and the prior distributions in the CRM method. According to the results of these comparisons, different fictions in the scenarios have been very effective in producing the selection probability of the right MTD of these methods. In the most scenarios, the methods selected the MTD correctly. In the simulation study of the final story, the changes in the results were effective in selecting the correct MTD when using different model structures and the prior distributions in the CRM method. The selection probabilities of the correct MTD obtained from some model structures were lower than the other model structures.

## 5.2 Conclusion

The phase I trials are the fundamental of drug development as they bring proposed designs to initial clinical testing. The dose escalation methods are very important at a selection of the MTD. In this thesis, we neither support nor denigrate all designs, but examine the properties and extensive simulation results of seven different methods. In addition, we presented the simulation results of the different model structures and prior distributions in the CRM design.

Overall, twenty different scenarios and three different stories were considered in the first and second part and the model-based dose response designs such as CRM, BMA-CRM mTPI and mTPI -2 reached the MTD with the similar selection percentages. In almost all simulations, the 3+3 design had the lowest selection percentage of the MTD. In the fourth scenario of the second story, where the true toxicity rate was changed very slightly, selected wrong dose level as the MTD in all designs. Therefore, simulation result where the toxicity rates changed very slightly was the most dramatic one among the other simulation results because the selection percentage of the MTD for the other simulations was correct. Thus, all designs should be checked before applying in a case when the true toxicity rate is changed very slightly.

In our simulation study, the BMA-CRM, CRM, mTPI and mTPI -2 performed the best results compared to other designs. The BMA-CRM design requires multiple skeletons in order to cover different scenarios. The BMA-CRM performs well if one of the skeletons corresponds to the true toxicity probabilities. In the scenarios, where the toxicity probability started at very low dose level and then increased slightly until the MTD (MTD is the last dose level), the BMA-CRM performed much better than the other designs. This design should be considered if the researcher has a similar scenario. When the number of the patients decreased from 30 to 21 in the second story, other methods appeared to be the best design for the selection of the MTD. For example, when the true toxicity rate started as the MTD then increased gradually to the high dose level, the BOIN selected the MTD with highest probability selection. Hansen et al. [81] found that the 3+3 design is appropriate when the toxicity of a drug is uncertain or a narrow. However, the 3+3 design appeared to be the worst design in our study. Our findings for the BMA-CRM design are similar to [58]. The BMA-CRM would perform well if one of the skeletons is similar to the true toxicity rate. We found similar results

to Paoletti et al. [82], claiming that the model-based methods outperform the 3+3 design in terms of selecting correct dose level as the MTD.

Overall, the CRM had the best selection percentage of the MTD among the designs. However, in some cases, the BMA-CRM, mTPI and mTPI-2 produced better results than the CRM. In contrast, even though the traditional 3+3 design is very simple and easy to apply, it performed the worst results among the other designs.

In general, the model-based designs produced better results. However, continual modeling by a professional is necessary. These designs can be complex for non-statisticians and might need statistical support [83]. In contrast, the 3+3 design does not require modeling and it offers protective dose escalation for candidate drugs, but the patients may be treated at sub-therapeutic doses and it may not be appropriate for molecularly targeted agents.

In the simulation studies for the third story, it was examined how the selection probabilities changed and the performance against the other methods when different model structures and the prior distributions of the CRM were used. Overall, in designs where the model structure is hyperbolic tangent and prior distribution is uniform, the CRM calculated the selection probability of the correct MTD higher. On the other hand, in designs where the model structure is logit and prior distribution is lognormal, the CRM calculated the selection probability of the correct MTD lower than the other CRM designs. In addition, The BMA-CRM produced very good results if the difference between the correct MTD dose and the previous-subsequent dose is greater. Considering these results, careful selection of the dose response methods to be selected in the application of the statistical method of the clinical trials to be performed is suggested.

In conclusion, more reliable and applicable results for the phase I dose finding trials are produced by the BMA-CRM and the CRM, when the model structure and the prior distributions are different, in our study. However, using mTPI and mTPI-2 designs can produce better results in the case where the target toxicity of the trial is not included in the study. As a result, the model-based designs performed much better than the rule-based designs.

## 5.3 Further Research

There are many possibilities for further research within the Bayesian approach. The flexibility of the Bayesian methods is more than frequentist methods, but data analysis should be performed after each patient has been treated and this can be challenging. In addition, selection of the prior distribution can be difficult. The historical data at which a prior distribution is modeled may not always be included in the trial. The prior distribution that has been chosen can be very informative and it can lead to inaccurate results in the new treatment. Calculations can be cumbersome for larger experiments and increase the chances of making erroneous decisions. To overcome this limitation, a hybrid method that deals with historical data in the Bayesian approach would be very interesting to investigate in the future.

## REFERENCES


[1] Friedman, L.M. Furberg, C.D. and DeMets, D.L., (2010). Basic study design, ed. Fundamentals of clinical trials. Springer, 67-96.

[2] Piantadosi, S., (2017). Clinical trials: a methodologic perspective: John Wiley & Sons.

[3] Spilker, B., (1992). "Guide to clinical trials".

[4] Yin, G., (2013). Clinical trial design: Bayesian and frequentist adaptive methods: John Wiley & Sons.

[5] Spodick, D.H., (1982). "The randomized controlled clinical trial: scientific and ethical bases", The American journal of medicine, 73: 420-425.

[6] Lee, J.J. Gu, X. and Liu, S., (2010). "Bayesian adaptive randomization designs for targeted agent development", Clinical Trials, 7: 584-596.

[7] Guidance, F.D., (2010). "Adaptive design clinical trials for drugs and biologics", Biotechnol Law Rep, 29: 173.

[8] Nelson, N.J., (2010). Adaptive clinical trial design: has its time come?, ed^eds.: Oxford University Press.

[9] Barker, A. Sigman, C. Kelloff, G. Hylton, N. Berry, D. and Esserman, L., (2009). "I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy", Clinical Pharmacology & Therapeutics, 86: 97-100.

[10] DiMasi, J.A. Hansen, R.W. and Grabowski, H.G., (2003). "The price of innovation: new estimates of drug development costs", Journal of health economics, 22: 151-185.

[11] Trouiller, P. Olliaro, P. Torreele, E. Orbinski, J. Laing, R. and Ford, N., (2002). "Drug development for neglected diseases: a deficient market and a public-health policy failure", The Lancet, 359: 2188-2194.

[12] Peck, C.C. Barr, W.H. Benet, L.Z. Collins, J. Desjardins, R.E. Furst, D.E. Harter, J.G. Levy, G. Ludden, T. and Rodman, J.H., (1992). "Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development", Journal of pharmaceutical sciences, 81: 605-610.

[13] Marchetti, S. and Schellens, J., (2007). "The impact of FDA and EMEA guidelines on drug development in relation to Phase 0 trials", British Journal of Cancer, 97: 577.

[14] Lesko, L.J. and Atkinson Jr, A., (2001). "Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria,

validation, strategies", Annual review of pharmacology and toxicology, 41: 347-366.

[15] Adams, C.P. and Brantner, V.V., (2006). "Estimating the cost of new drug development: is it really $802 million?", Health affairs, 25: 420-428.

[16] Jorgensen, W.L., (2004). "The many roles of computation in drug discovery", Science, 303: 1813-1818.

[17] Li, J.W.-H. and Vederas, J.C., (2009). "Drug discovery and natural products: end of an era or an endless frontier?", Science, 325: 161-165.

[18] O'Quigley, J. and Chevret, S., (1991). "Methods for dose finding studies in cancer clinical trials: a review and results of a Monte Carlo study", Statistics in medicine, 10: 1647-1664.

[19] Meinert, C.L., (2012). ClinicalTrials: Design, Conduct and Analysis: Oxford University Press.

[20] Thomann, M.A., (2015). The flexible bivariate continual reassessment method, ed^eds. Biostatistics. ProQuest Dissertations Publishing: The University of Iowa, 171.

[21] Julious, S.A. Machin, D. and Tan, S.B., (2010). An introduction to statistics in early phase trials: John Wiley.

[22] Bornkamp, B. Bretz, F. Dmitrienko, A. Enas, G. Gaydos, B. Hsu, C.H. Konig, F. Krams, M. Liu, Q. Neuenschwander, B. Parke, T. Pinheiro, J. Roy, A. Sax, R. and Shen, F., (2007). "Innovative approaches for designing and analyzing adaptive dose-ranging trials", Journal of Biopharmaceutical Statistics, 17: 965-995.

[23] Jennison, C. and Turnbull, B.W., (2001). "On group sequential tests for data in unequally sized groups and with unknown variance", Journal of Statistical Planning and Inference, 96: 263-288.

[24] Feng, H.B. Shao, J. and Chow, S.C., (2007). "Adaptive group sequential test for clinical trials with changing patient population", Journal of Biopharmaceutical Statistics, 17: 1227-1238.

[25] Akanuwe, J. Goyder, E. O'Hara, R. and Julious, S., (2010). "Exploring the Impact of the Quality and Outcomes Framework on the Quality of Diabetes Care and Health Care Inequalities in England", Journal of Epidemiology and Community Health, 64: A35-A35.

[26] Strimbu, K. and Tavel, J.A., (2010). "What are biomarkers?", Current Opinion in Hiv and Aids, 5: 463-466.

[27] Lachin, J.M., (1981). "Introduction to sample size determination and power analysis for clinical trials", Controlled clinical trials, 2: 93-113.

[28] Chow, S.C. Chang, M. and Pong, A., (2005). "Statistical consideration of adaptive methods in clinical development", Journal of Biopharmaceutical Statistics, 15: 575-591.

[29] Berry, D.A., (2006). "Bayesian clinical trials", Nature reviews. Drug discovery, 5: 27.

[30] Karanicolas, P.J. Farrokhyar, F. and Bhandari, M., (2010). "Blinding: Who, what, when, why, how?", Canadian journal of surgery, 53: 345.

[31] Wang, D. and Bakhai, A., (2006). Clinical trials: a practical guide to design, analysis, and reporting: Remedica.

[32] Pocock, S.J., (2013). Clinical trials: a practical approach: John Wiley & Sons.

[33] Chow, S.-C. and Liu, J.-P., (2008). Design and analysis of clinical trials: concepts and methodologies: John Wiley & Sons.

[34] Wallenstein, S. and Fisher, A.C., (1977). "The analysis of the two-period repeated measurements crossover design with application to clinical trials", Biometrics: 261-269.

[35] Brown Jr, B.W., (1980). "The crossover experiment for clinical trials", Biometrics: 69-79.

[36] Bernard, G.R. Artigas, A. Brigham, K.L. Carlet, J. Falke, K. Hudson, L. Lamy, M. Legall, J.R. Morris, A. and Spragg, R., (1994). "The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination", American journal of respiratory and critical care medicine, 149: 818-824.

[37] Peto, R. Pike, M. Armitage, P. Breslow, N.E. Cox, D. Howard, S. Mantel, N. McPherson, K. Peto, J. and Smith, P., (1977). "Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples", British Journal of Cancer, 35: 1.

[38] Sheiner, L.B. Stanski, D.R. Vozeh, S. Miller, R.D. and Ham, J., (1979). "Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to d-tubocurarine", Clinical Pharmacology & Therapeutics, 25: 358-371.

[39] Rosenbaum, S.E., (2016). Basic pharmacokinetics and pharmacodynamics: An integrated textbook and computer simulations: John Wiley & Sons.

[40] Chow, S.-C. and Chang, M., (2011). Adaptive design methods in clinical trials: CRC press.

[41] Hill, A.V., (1910). "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves", J Physiol (Lond), 40: 4-7.

[42] Chen, Z.J. Tighiouart, M. and Kowalski, J., (2012). "Dose escalation with overdose control using a quasi-continuous toxicity score in cancer Phase I clinical trials", Contemporary Clinical Trials, 33: 949-958.

[43] Vandemeulebroecke, M., (2008). "Group sequential and adaptive designs - A review of basic concepts and points of discussion", Biometrical Journal, 50: 541-557.

[44] Babb, J. Rogatko, A. and Zacks, S., (1998). "Cancer phase I clinical trials: efficient dose escalation with overdose control", Statistics in medicine, 17: 1103-1120.

[45] Senn, S.S., (2008). Statistical issues in drug development: John Wiley & Sons.

[46] Vandemeulebroecke, M., (2008). "Group sequential and adaptive designs–a review of basic concepts and points of discussion", Biometrical Journal, 50: 541-557.

[47] Gallo, P. Chuang-Stein, C. Dragalin, V. Gaydos, B. Krams, M. and Pinheiro, J., (2006). "Adaptive designs in clinical drug development—an executive summary of the PhRMA working group", Journal of Biopharmaceutical Statistics, 16: 275-283.

[48] Ji, Y. and Wang, S.-J., (2013). "Modified toxicity probability interval design: a safer and more reliable method than the 3+ 3 design for practical phase I trials", Journal of Clinical Oncology, 31: 1785-1791.

[49] Le Tourneau, C. Lee, J.J. and Siu, L.L., (2009). "Dose escalation methods in phase I cancer clinical trials", JNCI: Journal of the National Cancer Institute, 101: 708-720.

[50] James, G.D. Symeonides, S.N. Marshall, J. Young, J. and Clack, G., (2016). "Continual reassessment method for dose escalation clinical trials in oncology: a comparison of prior skeleton approaches using AZD3514 data", BMC cancer, 16: 703.

[51] Cater, S., (1972). "Study design principles for clinical evaluation of new drugs as developed by the chemotherapy program of the National Cancer Institute", The design of clinical trials in cancer therapy: 242-289.

[52] Le Tourneau, C. Gan, H.K. Razak, A.R. and Paoletti, X., (2012). "Efficiency of new dose escalation designs in dose-finding phase I trials of molecularly targeted agents", PloS one, 7: e51039.

[53] Storer, B.E., (1989). "Design and analysis of phase I clinical trials", Biometrics: 925-937.

[54] Wheeler, G.M. Sweeting, M.J. and Mander, A.P., (2016). "AplusB: A Web Application for Investigating A+ B Designs for Phase I Cancer Clinical Trials", PloS one, 11: e0159026.

[55] O'Quigley, J. Pepe, M. and Fisher, L., (1990). "Continual reassessment method: a practical design for phase 1 clinical trials in cancer", Biometrics: 33-48.

[56] Goodman, S.N. Zahurak, M.L. and Piantadosi, S., (1995). "Some practical improvements in the continual reassessment method for phase I studies", Statistics in medicine, 14: 1149-1161.

[57] Garrett-Mayer, E., (2006). "The continual reassessment method for dose-finding studies: a tutorial", Clinical Trials, 3: 57-71.

[58] Yin, G. and Yuan, Y., (2009). "Bayesian model averaging continual reassessment method in phase I clinical trials", Journal of the American Statistical Association, 104: 954-968.

[59] Cheung, Y.K. and Chappell, R., (2000). "Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities", Biometrics, 56: 1177-1182.

[60] Faries, D., (1994). "Practical modifications of the continual reassessment method for phase I cancer clinical trials", Journal of Biopharmaceutical Statistics, 4: 147-164.

[61] Korn, E.L. Midthune, D. Chen, T.T. Rubinstein, L.V. Christian, M.C. and Simon, R.M., (1994). "A comparison of two phase I trial designs", Statistics in medicine, 13: 1799-1806.

[62] Gasparini, M. and Eisele, J., (2000). "A Curve-Free Method for Phase I Clinical Trials", Biometrics, 56: 609-615.

[63] Piantadosi, S. Fisher, J. and Grossman, S., (1998). "Practical implementation of a modified continual reassessment method for dose-finding trials", Cancer chemotherapy and pharmacology, 41: 429-436.

[64] Ishizuka, N. and Ohashi, Y., (2001). "The continual reassessment method and its applications: a Bayesian methodology for phase I cancer clinical trials", Statistics in medicine, 20: 2661-2681.

[65] Sweeting, M. Mander, A. and Sabin, T., (2013). "Bcrm: Bayesian continual reassessment method designs for phase I dose-finding trials", Journal of Statistical Software, 54: 1-26.

[66] Yuan, Z. Chappell, R. and Bailey, H., (2007). "The Continual Reassessment Method for Multiple Toxicity Grades: A Bayesian Quasi-Likelihood Approach", Biometrics, 63: 173-179.

[67] Hoeting, J.A. Madigan, D. Raftery, A.E. and Volinsky, C.T., (1999). "Bayesian model averaging: a tutorial", Statistical science: 382-401.

[68] Madigan, D. and Raftery, A.E., (1994). "Model selection and accounting for model uncertainty in graphical models using Occam's window", Journal of the American Statistical Association, 89: 1535-1546.

[69] Yuan, Y. Hess, K.R. Hilsenbeck, S.G. and Gilbert, M.R., (2016). "Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials", Clinical Cancer Research: clincanres. 0592.2016.

[70] Liu, S. and Yuan, Y., (2015). "Bayesian optimal interval designs for phase I clinical trials", Journal of the Royal Statistical Society: Series C (Applied Statistics), 64: 507-523.

[71] Chu, Y. Pan, H. and Yuan, Y., (2016). "Adaptive dose modification for phase I clinical trials", Statistics in medicine, 35: 3497-3508.

[72] Ji, Y. Liu, P. Li, Y. and Nebiyou Bekele, B., (2010). "A modified toxicity probability interval method for dose-finding trials", Clinical Trials, 7: 653-663.

[73] Guo, W. Wang, S.-J. Yang, S. Lynn, H. and Ji, Y., (2017). "A Bayesian interval dose-finding design addressingOckham's razor: mTPI-2", Contemporary Clinical Trials, 58: 23-33.

[74] Yang, S. Wang, S.-J. and Ji, Y., (2015). "An integrated dose-finding tool for phase I trials in oncology", Contemporary Clinical Trials, 45: 426-434.

[75] Thorburn, W.M., (1918). "The myth of Occam's razor", Mind, 27: 345-353.

[76] Jefferys, W.H. and Berger, J.O., (1992). "Ockham's razor and Bayesian analysis", American Scientist, 80: 64-72.

[77] Myung, I.J. and Pitt, M.A., (1997). "Applying Occam’s razor in modeling cognition: A Bayesian approach", Psychonomic Bulletin & Review, 4: 79-95.

[78] A Tolerability and Pharmacokinetics Study of SHR6390 in Advanced Melanoma Patients, url.: https://ClinicalTrials.gov/show/NCT02671513.

[79] Reverse Genetic H9N2 Influenza Vaccine Study in Adults, url: https://ClinicalTrials.gov/show/NCT01320696.

[80] Phase I Trial of Tanibirumab in Advanced or Metastatic Cancer, url: https://ClinicalTrials.gov/show/NCT01660360.

[81] Hansen, A.R. Graham, D.M. Pond, G.R. and Siu, L.L., (2014). "Phase 1 trial design: is 3+ 3 the best?", Cancer Control, 21: 200-208.

[82] Paoletti, X. Ezzalfani, M. and Le Tourneau, C., (2015). "Statistical controversies in clinical research: requiem for the 3+ 3 design for phase I trials", Annals of Oncology, 26: 1808-1812.

[83] Wong, K.M. Capasso, A. and Eckhardt, S.G., (2016). "The changing landscape of phase I trials in oncology", Nature reviews. Clinical oncology, 13: 106.

# APPENDIX-A

## COMPARISON OF CLINICAL TRIALS PHASES

Table A.1 Comparison of the clinical trials phases

| Phase | Primary Goal | Factors | Dose | Monitoring of patients | Design features | Duration | Population |
|---|---|---|---|---|---|---|---|
| **Phase I** | Understand the metabolic actions and the MTD | -Bioavailability<br>-Metabolism<br>-Pharmacokinetics<br>-Pharmacodynamics | Sub therapeutic | Clinicians | -Unblinded<br>-Uncontrolled | Up to 1 month | 20-50 people |
| **Phase II** | Efficacy and safety of drugs | -Bioavailability<br>-Patient Safety<br>-Pharmacokinetics<br>-Pharmacodynamics<br>-Efficacy at various doses | Therapeutic dose | Clinicians | -Placebo controlled comparisons<br>-Active controlled comparison | Several months | 200-300 people |
| **Phase III** | Efficacy and safety of drugs | -Dosage intervals<br>-Risk-benefit information<br>- Drug interactions | Therapeutic dose | Clinicians and personal physicians | -Randomized<br>-Controlled | Several years | 100-1000 people |
| **Phase IV** | Monitoring safety in large population | -Epidemiological data<br>- Pharmacoeconomics | Therapeutic dose | personal physicians | -Observational<br>-Uncontrolled | Ongoing (Depends on the FD approval) | 1000s people |

The detailed comparison of the clinical trials phases is given in Table A.1.

# APPENDIX-B

## A+B DESIGN SIMULATION RESULTS

This appendix presents the simulation results that were performed for different variations of A+B design for story 1 and story 2. The comparison results of the simulation study for story 1 is given in Table B.1-B.3 and the simulation results of the different variations of A+B design for story 2 is given in Table B.4-B.6.

Table B.1 Comparison of different variations of A+B design in scenario I and II

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg /day** | **75 mg/day** | **100 mg/day** | **125 mg/day** | **150 mg/day** | **175 mg/day** |
| | **Method** | True toxicity rate | 0.02 | 0.06 | 0.12 | 0.2 | **0.3** | 0.4 |
| **Scenario1** | **3+3** | Probability selection | 0.054 | 0.215 | 0.265 | 0.395 | **0.060** | 0.005 |
| | **2+2** | Probability selection | 0.17 | 0.18 | 0.19 | 0.19 | **0.16** | 0.11 |
| | **2+3** | Probability selection | 0.17 | 0.18 | 0.20 | 0.19 | **0.16** | 0.10 |
| | **3+2** | Probability selection | 0.20 | 0.21 | 0.21 | 0.18 | **0.13** | 0.07 |
| | **2+4** | Probability selection | 0.16 | 0.18 | 0.20 | 0.20 | **0.16** | 0.09 |
| | **4+2** | Probability selection | 0.22 | 0.23 | 0.22 | 0.18 | **0.11** | 0.04 |
| | **4+4** | Probability selection | 0.22 | 0.24 | 0.24 | 0.18 | **0.09** | 0.03 |
| | **Method** | True toxicity rate | 0.05 | 0.15 | **0.30** | 0.46 | 0.61 | 0.73 |
| **Scenario2** | **3+3** | Probability selection | 0.217 | 0.268 | **0.428** | 0.057 | 0.004 | 0.000 |
| | **2+2** | Probability selection | 0.090 | 0.271 | **0.350** | 0.212 | 0.050 | 0.011 |
| | **2+3** | Probability selection | 0.122 | 0.320 | **0.352** | 0.170 | 0.031 | 0.000 |
| | **3+2** | Probability selection | 0.153 | 0.371 | **0.347** | 0.122 | 0.010 | 0.000 |
| | **2+4** | Probability selection | 0.140 | 0.342 | **0.333** | 0.146 | 0.031 | 0.000 |
| | **4+2** | Probability selection | 0.202 | 0.436 | **0.288** | 0.050 | 0.000 | 0.000 |
| | **4+4** | Probability selection | 0.271 | 0.455 | **0.210** | 0.022 | 0.000 | 0.000 |

Table B.2 Comparison of different variations of A+B design in scenario III and IV

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/day** | **75 mg/day** | **100 mg/day** | **125 mg/day** | **150 mg/day** | **175 mg/day** |
| | **Method** | True toxicity rate | 0.02 | 0.07 | 0.16 | **0.3** | 0.44 | 0.57 |
| **Scenario3** | **3+3** | Probability selection | 0.054 | 0.215 | 0.265 | **0.395** | 0.060 | 0.005 |
| | **2+2** | Probability selection | 0.021 | 0.107 | 0.270 | **0.322** | 0.214 | 0.080 |
| | **2+3** | Probability selection | 0.031 | 0.131 | 0.313 | **0.325** | 0.172 | 0.050 |
| | **3+2** | Probability selection | 0.040 | 0.160 | 0.351 | **0.319** | 0.124 | 0.020 |
| | **2+4** | Probability selection | 0.043 | 0.154 | 0.336 | **0.300** | 0.141 | 0.033 |
| | **4+2** | Probability selection | 0.061 | 0.225 | 0.403 | **0.264** | 0.066 | 0.000 |
| | **4+4** | Probability selection | 0.082 | 0.282 | 0.416 | **0.180** | 0.030 | 0.000 |
| | **Method** | True toxicity rate | 0.05 | 0.1 | 0.15 | 0.2 | **0.3** | 0.4 |
| **Scenario4** | **3+3** | Probability selection | 0.105 | 0.165 | 0.208 | 0.253 | **0.297** | 0.078 |
| | **2+2** | Probability selection | 0.19 | 0.20 | 0.23 | 0.19 | **0.13** | 0.06 |
| | **2+3** | Probability selection | 0.18 | 0.20 | 0.25 | 0.20 | **0.12** | 0.05 |
| | **3+2** | Probability selection | 0.22 | 0.23 | 0.26 | 0.18 | **0.08** | 0.03 |
| | **2+4** | Probability selection | 0.17 | 0.20 | 0.27 | 0.21 | **0.12** | 0.04 |
| | **4+2** | Probability selection | 0.25 | 0.26 | 0.27 | 0.15 | **0.06** | 0.01 |
| | **4+4** | Probability selection | 0.271 | 0.455 | 0.210 | 0.022 | **0.000** | 0.000 |

Table B.3 Comparison of different variations of A+B design in scenario V and VI

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **50mg/day** | **75 mg/day** | **100 mg/day** | **125 mg/day** | **150 mg/day** | **175 mg/day** |
| | **Method** | True toxicity rate | 0.18 | **0.3** | 0.42 | 0.53 | 0.64 | 0.72 |
| **Scenario5** | **3+3** | Probability selection | 0.311 | **0.425** | 0.171 | 0.095 | 0 | 0 |
| | **2+2** | Probability selection | 0.39 | **0.32** | 0.19 | 0.08 | 0.02 | 0.00 |
| | **2+3** | Probability selection | 0.41 | **0.33** | 0.18 | 0.07 | 0.01 | 0.00 |
| | **3+2** | Probability selection | 0.50 | **0.32** | 0.14 | 0.04 | 0.01 | 0.00 |
| | **2+4** | Probability selection | 0.42 | **0.33** | 0.17 | 0.06 | 0.01 | 0.00 |
| | **4+2** | Probability selection | 0.58 | **0.30** | 0.10 | 0.02 | 0.00 | 0.00 |
| | **4+4** | Probability selection | 0.63 | **0.29** | 0.07 | 0.01 | 0.00 | 0.00 |
| | **Method** | True toxicity rate | 0.01 | **0.3** | 0.55 | 0.65 | 0.8 | 0.95 |
| **Scenario6** | **3+3** | Probability selection | 0.34 | **0.47** | 0.17 | 0.02 | 0.00 | 0.00 |
| | **2+2** | Probability selection | 0.29 | **0.40** | 0.25 | 0.06 | 0.01 | 0.00 |
| | **2+3** | Probability selection | 0.28 | **0.43** | 0.24 | 0.05 | 0.01 | 0.00 |
| | **3+2** | Probability selection | 0.35 | **0.44** | 0.19 | 0.02 | 0.00 | 0.00 |
| | **2+4** | Probability selection | 0.36 | **0.35** | 0.24 | 0.04 | 0.00 | 0.00 |
| | **4+2** | Probability selection | 0.49 | **0.36** | 0.14 | 0.01 | 0.00 | 0.00 |
| | **4+4** | Probability selection | 0.48 | **0.40** | 0.11 | 0.00 | 0.00 | 0.00 |

Table B.4 Comparison of different variations of A+B design in scenario I and II

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | 0.15 | 0.2 | **0.25** | 0.3 | 0.35 | 0.4 |
| **Scenario1** | **3+3** | Probability selection | 0.133 | 0. 165 | **0.283** | 0.203 | 0.134 | 0.117 |
| | **2+2** | Probability selection | 0.31 | 0.25 | **0.18** | 0.13 | 0.08 | 0.05 |
| | **2+3** | Probability selection | 0.33 | 0.26 | **0.19** | 0.12 | 0.07 | 0.04 |
| | **3+2** | Probability selection | 0.40 | 0.27 | **0.17** | 0.09 | 0.05 | 0.02 |
| | **2+4** | Probability selection | 0.34 | 0.27 | **0.19** | 0.12 | 0.06 | 0.03 |
| | **4+2** | Probability selection | 0.48 | 0.28 | **0.15** | 0.07 | 0.02 | 0.01 |
| | **4+4** | Probability selection | 0.53 | 0.28 | **0.13** | 0.05 | 0.01 | 0.00 |
| | **Method** | True toxicity rate | 0.01 | 0.02 | 0.03 | 0.04 | 0.1 | **0.25** |
| **Scenario2** | **3+3** | Probability selection | 0.004 | 0.011 | 0.014 | 0.091 | 0.367 | **0.51** |
| | **2+2** | Probability selection | 0.16 | 0.16 | 0.16 | 0.16 | 0.17 | **0.19** |
| | **2+3** | Probability selection | 0.15 | 0.15 | 0.16 | 0.16 | 0.18 | **0.20** |
| | **3+2** | Probability selection | 0.16 | 0.16 | 0.16 | 0.16 | 0.17 | **0.17** |
| | **2+4** | Probability selection | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | **0.16** |
| | **4+2** | Probability selection | 0.15 | 0.15 | 0.15 | 0.16 | 0.18 | **0.21** |
| | **4+4** | Probability selection | 0.16 | 0.17 | 0.17 | 0.17 | 0.18 | **0.16** |

Table B.5 Comparison of different variations of A+B design in scenario III and IV

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | **0.25** | 0.35 | 0.4 | 0.5 | 0.6 | 0.7 |
| **Scenario3** | **3+3** | Probability selection | **0.352** | 0.327 | 0.291 | 0.025 | 0.001 | 0 |
| | **2+2** | Probability selection | **0.38** | 0.39 | 0.14 | 0.06 | 0.02 | 0.00 |
| | **2+3** | Probability selection | **0.31** | 0.50 | 0.13 | 0.05 | 0.01 | 0.00 |
| | **3+2** | Probability selection | **0.40** | 0.47 | 0.09 | 0.03 | 0.00 | 0.00 |
| | **2+4** | Probability selection | **0.42** | 0.40 | 0.12 | 0.04 | 0.01 | 0.00 |
| | **4+2** | Probability selection | **0.49** | 0.44 | 0.06 | 0.01 | 0.00 | 0.00 |
| | **4+4** | Probability selection | **0.45** | 0.51 | 0.04 | 0.00 | 0.00 | 0.00 |
| | **Method** | True toxicity rate | 0.19 | 0.21 | 0.23 | **0.25** | 0.27 | 0.29 |
| **Scenario4** | **3+3** | Probability selection | 0.339 | 0.291 | 0.240 | **0.079** | 0.046 | 0.005 |
| | **2+2** | Probability selection | 0.35 | 0.23 | 0.16 | **0.12** | 0.08 | 0.06 |
| | **2+3** | Probability selection | 0.38 | 0.24 | 0.16 | **0.11** | 0.07 | 0.05 |
| | **3+2** | Probability selection | 0.45 | 0.24 | 0.14 | **0.08** | 0.05 | 0.03 |
| | **2+4** | Probability selection | 0.39 | 0.25 | 0.16 | **0.10** | 0.06 | 0.04 |
| | **4+2** | Probability selection | 0.54 | 0.24 | 0.12 | **0.06** | 0.03 | 0.01 |
| | **4+4** | Probability selection | 0.271 | 0.455 | 0.210 | 0.022 | 0.000 | 0.000 |

Table B.6 Comparison of different variations of A+B design in scenario V and VI

| | | | Dose Levels | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **3,75mg/d** | **7,5mg/d** | **15mg/d** | **30mg/d** | **45mg/d** | **60mg/d** |
| | **Method** | True toxicity rate | 0.01 | 0.1 | 0.2 | **0.25** | 0.3 | 0.35 |
| **Scenario5** | **3+3** | Probability selection | 0.093 | 0.184 | 0.242 | **0.297** | 0.109 | 0.074 |
| | **2+2** | Probability selection | 0.18 | 0.21 | 0.22 | **0.17** | 0.13 | 0.08 |
| | **2+3** | Probability selection | 0.18 | 0.22 | 0.23 | **0.17** | 0.12 | 0.07 |
| | **3+2** | Probability selection | 0.22 | 0.25 | 0.23 | **0.15** | 0.09 | 0.05 |
| | **2+4** | Probability selection | 0.17 | 0.23 | 0.24 | **0.18** | 0.11 | 0.07 |
| | **4+2** | Probability selection | 0.25 | 0.28 | 0.24 | **0.14** | 0.06 | 0.03 |
| | **4+4** | Probability selection | 0.25 | 0.31 | 0.26 | **0.12** | 0.05 | 0.01 |
| | **Method** | True toxicity rate | 0.02 | 0.04 | 0.08 | 0.12 | 0.17 | **0.25** |
| **Scenario6** | **3+3** | Probability selection | 0.019 | 0.071 | 0.134 | 0.193 | 0.232 | **0.351** |
| | **2+2** | Probability selection | 0.16 | 0.17 | 0.17 | 0.17 | 0.17 | **0.15** |
| | **2+3** | Probability selection | 0.16 | 0.17 | 0.18 | 0.18 | 0.17 | **0.15** |
| | **3+2** | Probability selection | 0.18 | 0.18 | 0.18 | 0.17 | 0.16 | **0.13** |
| | **2+4** | Probability selection | 0.15 | 0.16 | 0.18 | 0.18 | 0.17 | **0.15** |
| | **4+2** | Probability selection | 0.19 | 0.19 | 0.19 | 0.17 | 0.14 | **0.10** |
| | **4+4** | Probability selection | 0.19 | 0.20 | 0.20 | 0.18 | 0.14 | **0.09** |

# CURRICULUM VITAE

## PERSONAL INFORMATION

**Name Surname** : Efehan ULAŞ

**Date of birth and place** : 01.01.1989 Mamak

**Foreign Languages** : English

**E-mail** : efehanulas@karatekin.edu.tr

## EDUCATION

| Degree | Department | University | Date of Graduation |
|---|---|---|---|
| Doctorate | Statistics | Yıldız Technical University | 03.11.2017 |
| Master | Statistics | The University of Sheffield | 05.11.2012 |
| Undergraduate | Statistics | Muğla Sıtkı Koçman University | 16.06.2010 |
| High School | Science | Sokullu Mehmetpaşa High School | 12.06.2005 |

## WORK EXPERIENCE

| Year | Corporation/Institute | Enrollment |
|---|---|---|
| 2013 | Cankiri Karatekin University | 2013 |

## PUBLISHMENTS

### Papers

1. Ulas E, Karaman F (2018), ” A comprehensive comparison of phase I dose escalation methods”, Kuwait Journal of Science , Vol 45(4), 2018 (Accepted)

2. Edis S, Ulas E (2017), ” Using Bayesian Network to Predict The Watershed Land Use Type of Çankırı Acıçay-Tatlıçay”, Turkish Journal of Foresty (Accepted)

3. Koc T, Ulas E, Koc H, Keskin B (2017) “Anaylsis of Social Media Usage Habits Among the People of Samsun”, Journal of Basic and Applied Research International, 22(3): 93-98, 2017

4. Keskin B, Ulas E, Koc H (2016) “Financial Performance Analysis by Using TOPSIS and ELECTRE Methods: A Research on Turkish Construction Sector Companies”, International Journal of Sciences: Basic and Applied Research, 30(5): 156-164, 2016

5. Ulas E, Keskin B (2015) ”Performance Evaluation and Ranking of Turkish Banking Sector”,Procedia Economics and Finance (2015) pp. 297-307

### Conference Papers

1. Edis S, Ulas E (2017), “Determination of Watershed Land Use Type with Bayesian Network in Semi Arid Region”, International Conference On Agriculture, Forest, Food Sciences And Technologies

2. Edis S, Ulas E (2017), “Trend Analyses of Standardized Precipitation Index in Konya Endorheic Basin, Turkey”, International Conference On Agriculture, Forest, Food Sciences And Technologies

3. Keskin B, Ulas E (2017), “Investigation of the Relationship Between Brand Value and Performances of the Technology Companies”, International Congress on Management Economics and Business

4. Edis S, Ulas E (2017), “Desertification assesment using the standardized precipitation index in Cankiri, Turkey”, International Symposium on New Horizons in Forestry-ISFOR 2017, Isparta, Turkey

5. Edis S, Ulas E (2017), “Using machine learning methods in order to classify watershed land use type”, International Symposium on New Horizons in Forestry-ISFOR 2017, Isparta, Turkey

6. Koc T, Cengiz MA, Koc H, Ulas E (2015), “Determining factors that affects housing prices in Turkey by Bayesian network”, 12th Applied Statistics 2015 International Conference Bled

7. Koc H, Cengiz MA, Koc T, Ulas E (2015), “Determination of the profile of social media usage habits by the Latent Class Analysis”, 12th Applied Statistics

2015International Conference, Bled

8. Keskin B, Ulas E, Filiz E (2015), “Estimating Technical and Scale Efficiencies in Airline Industry: A Non-Parametric DEA Approach”, 12th Applied Statistics 2015 International Conference

9. Ulas E, Bas A, Aktan M (2014), “Analyzing the Metal Exchange Data with R, Multidisciplinary Academic Conference on Economics”, Management and Marketing, Prague

10. Keskin B, Ulas E (2014), “Measurement of Efficiency Using Data Envelopment Analysis: An Application in Airline Industry”, 20th Conference of the International Federation of Operational Research Societies, Barcelona

11. Ulas E , Keskin B (2014) “Forecasting the Temperature in Melbourne”, Australia, 9th International Statistics Days Symposium, Side

**Book Chapters**

1. Ulas E, Keskin B (2017), ” Is There a Relation between HDI and Economic Performances?”, Springer Prooceedings in Business and Economics (2017), doi. 10.1007/978- 3-319-49559-0

2. Keskin B, Ulas E (2017), ” Does Privatization Affect Airports Performance? A Comparative Analysis with AHP-TOPSIS and DEA”, Springer Prooceedings in Business and Economics (2017), doi. 10.1007/978-3-319-49559-0

**Projects**

1. Ulas E (2016), “Avrupa'da En Fazla Ziyaret Edilen Havalimanlarının Performanslarının Belirlenmesi”, Cankiri Karatekin University, Research Project

2. Ulas E(2016),”Yarı Kurak Havzalarda GIS ve Hidro-Morfolojik Tabanlı Tahmin Modellemesi;Çankırı Tatlıçay-Acıçay Örneklemi”,Cankiri Karatekin University, Research Project

**AWARDS**

1. TUBİTAK 2224-A Yurt Dışı Bilimsel Etkinliklere Katılma Desteği (2015)

2. YLSY, Milli Eğitim Bakanlığı Yurtdışı Yüksek Lisans Bursu (2010)