Novel Statistical Designs for Phase I clinical trials

by

Weijia Zhang

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Statistics
The University of Manitoba
Winnipeg

Abstract

A clinical trial is an experiment on human subjects designed to evaluate the safety and efficacy of a new drug or medical intervention. There are four phases of a clinical trial. Phase I trial is the first step to determine the maximum tolerated dose (MTD) to be used in the subsequent trials. Phase II trial examines the new drug's short-term efficacy based on the MTD identified in Phase I. It is also the proof-of-concept trial for Phase III. If the drug shows potential efficacy for the larger population, we can move onto a Phase III trial. Phase III trial is an expensive and long-term trial before the approval for marketing the new drug. It concerns about the long-term efficacy and safety of the new drug. If the new drug is approved for marketing, the Phase VI trial is the post-marketing surveillance trial that reports side effects of the new drug after marketing.

My Ph.D. research is focused on proposing and evaluating statistical designs of Phase I clinical trials. Currently there are three classic parametric designs available in the literature for Phase I clinical trials. The commonly used parametric design is the continual reassessment method (CRM). This method assumes a parametric statistical model to describe toxicity probability at each dose level. However, this parametric model has unknown parameters. These unknown parameters follow prior distributions under the Bayesian approach. After treating patients in the trial, we observed outcomes which are either toxic or nontoxic. Observations

of patients' outcomes are used to update posterior mean toxicity probabilities. The MTD is identified as the dose whose posterior mean toxicity probability is closest to the target toxicity probability, say 33%, after all patients in the trial are treated. The objective of a Phase I trial is to determine the MTD to be used in the subsequent clinical trials.

Three classic parametric models are normally used with the CRM, namely the power, logistic and hyperbolic tangent models. They use respectively the power, logistic and hyperbolic tangent functions. These functions are used to define the increasing relationship between toxicity probabilities and different dose levels.

In this thesis, we use the CRM with a new class of parametric functions. This class is based on the cumulative distribution function of the normal distribution. A major advantage is that we can choose different values of the mean and variance of the normal distribution to change the location and shape of the dose toxicity probability curve so we can flexibly model different shapes of the dose toxicity probability relationship. We conduct simulation studies and compare our new design with the existing designs, for one drug or the combination of two drugs. We investigate the performance of our new design when we assume that the variance is unknown, and the performance of the Bayesian model averaging CRM design. Finally we derive asymptotic statistical inference of the unknown parameter. We introduce some new performance criteria and compare different models based on "BEARS": Benchmark, Efficiency, Accuracy, Reliability, Safety.

In summary, Our designs performs well by choosing the appropriate values of α, β, the mean and variance in our model under criteria BEARS.

Acknowledgements

I would like to thank my thesis co-advisors Dr. Po Yang and Dr. Saman Muthukumarana for their patience and support during my thesis research. Their comments and guidance have been very helpful. I also thank my PhD Thesis Advisory Committee members, Dr. Saumen Mandal, Dr. Xiaofan Zheng, and the external examiner Dr. Dianliang Deng (University of Regina), for their constructive criticism.

I acknowledge financial supports from the Faculty of Graduate Studies, the Faculty of Science, the Department of Statistics, and from Dr. Yang's research grants and Dr. Muthukumarana's research grants.

Finally, I thank members of the Department of Statistics for their support and encouragement.

To My Parents

Contents

Abstract **i**

Acknowledgements **iii**

Dedication **iv**

List of abbreviations **viii**

1 Introduction **1**

1.1 Introduction 1

1.2 Nonparametric methods 4

1.2.1 3+3 Methods and extension 4

1.2.2 A+B Design 5

1.2.3 Accelerated titration design 6

1.3 Parametric methods 7

1.3.1 Continual reassessment method 7

1.3.2 Bayesian model averaging CRM 10

1.3.3 Escalation with overdose control 12

1.3.4 CRM with combination drugs 14

1.4 Summary 16

2 The CRM with one drug in Phase I trials 18

2.1 Introduction 18

2.2 A new design of CRM for one drug 19

2.3 Simulation results and discussion 26

2.4 Application 35

2.5 Conclusion 38

3 BMA-CRM 39

3.1 Introduction 39

3.2 Bayesian Model Averaging-CRM 40

3.3 Simulation Study 43

3.4 Conclusion 48

4 CRM without undue risk of toxicity 53

4.1 Introduction 53

4.2 The new CRM design 55

4.3 Simulation results and discussion 58

4.4 Conclusions 69

5 The CRM design with combination drugs 70

5.1 Introduction 70

5.2 Model-Based Designs for Combination Drugs 73

5.3 Possible dose-toxicity orders 75

5.4 Simulation study 79

5.5 Conclusion 84

6 Statistical Inference 93

6.1 Introduction 93

6.2 Maximum Likelihood Estimation 94

6.3 Numerical Illustration 100

6.4 Conclusion 100

7 Summary 102

Bibliography 104

List of abbreviations

- FDA: U.S. Food and Drug Administration
- MTD: maximum tolerated dose
- ATD: accelerated titration design
- BCD: biased coin dose-finding
- CRM: continual reassessment method
- DLT: dose limiting toxicity
- BMA: Bayesian model averaging
- BMA-CRM: Bayesian model averaging-continual reassessment method
- EWOC: escalation with overdose control
- BEARS: Benchmark, Efficiency, Accuracy, Reliability, Safety
- PCS: percentage of correct selection

Chapter 1

Introduction

1.1 Introduction

Clinical trials are designed experiments on human subjects and are typically characterized by the tension between collective and individual ethics. In a clinical trial, we hold dual responsibilities to current and future patients. Collective ethics requires that we, as scientists, conduct scientifically valid clinical research to advance medical knowledge so future patients benefit. Individual ethics dictates that we, as medical care providers, must ensure safety to each individual patient in the trial and provide the best possible treatment available based on current information.

Phase I clinical trial is the first stage of a drug tested in human beings before its approval of U.S. Food and Drug Administration (FDA). The overarching goal of a phase I clinical trial is to identify the maximum tolerated dose (MTD). Other goals include minimizing the overall risk of toxicity and treating as many patients

as possible at the therapeutic MTD. Phase I clinical trials are fundamentally important for the proof-of-concept phase II and confirmatory phase III clinical trials and play critical roles in cancer research and treatment (Weber et al., 2015), because we need to recommend the MTD to Phase II and Phase III clinical trials. If the MTD is over-estimated, patients may be exposed to a fatal treatment in following trials. If the MTD is under-estimated, the treatment may be ineffective. Trade-off between efficacy and toxicity of the new drug is our goal in all phases of clinical trials, but Phase I trial is only focused on the maximum tolerated dose that patients are not able to tolerate any more. Generally speaking, a Phase I clinical trial needs 15 to 30 patients (MDAnderson, Accessed 2019-05-13). Furthermore, the trial starts with a very low and safe dose. Patients are treated in cohorts, with a particular cohort size. The first cohort is treated at a starting dose. If a large number of patients in this cohort experience toxicity, the next cohort may be treated at a lower dose. Otherwise the next cohort may be given a higher dose. The dose with an estimated toxicity probability nearest to a pre-specified target dose limiting toxicity (DLT) rate is chosen as the MTD. We typically assume that toxicity and efficacy probabilities increase with increasing dose levels. Phase I clinical trials are also called dose finding trials. We expect to observe toxicity immediately for decision making. We face survived data when we need to determine whether or not to escalate or de-escalate the dose level for the next cohort, which the toxicity response is delayed.

As we know, the goal of a Phase I clinical trial is to identify the MTD for

later phase studies. Some important issues of this trial must be taken into account: ethical concerns, the starting dose, the speed of dose escalation, the sample size of patients, and the target toxicity probability. These issues have direct impact on the efficiency of Phase I trial design and the accuracy of the identified MTD.

Dose finding designs for phase I clinical trials are broadly classified into two groups: nonparametric (or algorithm-based) and parametric (or model-based). Such designs are often adaptive in the sense that selecting the next dose depends on the current dose and toxicity response. Nonparametric designs include the 3+3 method and its extensions (Storer, 1989), the accelerated titration design (Simon et al., 1997), the biased coin design (Durham et al., 1997), the group up-and-down design (Gezmu and Flournoy, 2006)), and the interval design (Ivanova et al., 2007). Nonparametric designs are easy to implement but not necessarily optimal, because they do not specifically target the MTD and there is no comprehensive understanding of the operating characteristic of the design.

A parametric design assumes an explicit statistical model of dose toxicity probability over the set of all possible dose levels. This approach is holistic because each assigned dose and the toxicity response are used to update the entire statistical relationship. The updated relationship is then used to identify a dose for the next patient. This model-based approach is introduced by O'Quigley et al. (1990) who propose the continual reassessment method (CRM). The CRM design specifies the dose toxicity probability by a parametric statistical model with

unknown parameters. The other parametric designs include the Escalation with Overdose Control (EWOC) (Babb et al., 1998), and Bayesian decision-theoretic approaches (Whitehead and Brunier, 1995).

1.2 Nonparametric methods

1.2.1 3+3 Methods and extension

The challenge of Phase I trial is to find the MTD subject to safety and efficiency. That means we need to avoid toxic doses higher than the MTD in treating patients and also ensure that the design is effective so that doses lower than the MTD are not frequently applied to patients.

The 3+3 design is a standard and classic method in finding the MTD with a targeted toxicity probability. Storer (1989) shows that in such a design, the toxicity probability is less than 33%. In a real clinical trial, the 3+3 design is widely used because of its easy implementation. We assume a pre-specified toxicity probability p_i which is increasing in dose level d_i. The cohort size is 3 and the first 3 patients are treated at the lowest starting dose level. The process of 3+3 design is described as follows:

(1) Treat three patients at the current dose level i.

(2) If no patient is toxic, escalate to the next higher dose level $i+1$, and go back to step (1).

(3) If one patient is toxic, then three more patients are treated at this same dose level i and we observe dose limiting toxicity (DLT) of the total of six patients. If only one of the group of six patients experiences toxicity, we move to the next higher dose level $i+1$. If two of six patients are toxic, the trial is terminated and the next lower dose level $i-1$ is considered as the MDT. If more than two patients of these six patients are toxic, then the current dose level i is higher than the MTD, and another three patients will enter the treatment and be treated at the next lower dose level $i-1$.

(4) If more than one patient is toxic at the lowest dose level, the trial will be terminated and is said to be an inconclusive trial.

Although the 3+3 design is convenient to use, it has some problems that may affect the estimation of dose-finding (O'Quigley and Chevret, 1991; O'Quigley and Shen, 1996; Yin, 2012). First, the observation data is only associated with the current dose level, and other dose levels are not considered. Second, the 3+3 design has poor statistical properties. Third, the 3+3 design only works for a trial in which the target toxicity probability is smaller than 33%.

1.2.2 A+B Design

Procedures of A+B design are described as follows:

(1) Suppose that A patients are treated at the dose level i, and we observe the outcome of toxicity.

(2) If less than C patients of the total of A patients are observed toxic, we escalate to the higher dose level $i + 1$.

(3) If the number of toxic patients is between C and D, we stay at the same dose level i where we will treat B more patients. If we observe that more than E of A+B patients are toxic, de-escalate to the next lower dose level $i - 1$, otherwise, we escalate to the next higher dose level $i + 1$.

(4) If the number of toxic patients are greater than D, de-escalate to the dose level $i - 1$.

As we know, the standard 3+3 design is a special case of A+B design when A and B are 3 and C, D and E are 1. Lin and Shih (2001) extend the 3+3 design to A+B design that is more general in practice. The significant difference from the 3+3 design is that the cohort size may not be three.

1.2.3 Accelerated titration design

In a classic Phase I design, to be safe, we start with the very low dose level to protect treated patients from toxicity, however, the determined MTD is usually higher than the starting dose. To speed up the trial, Simon et al. (1997) develop the accelerated titration design (ATD) instead of the traditional 3+3 design. The first stage of the ATD is to treat only one patient at each sequential dose level until observing toxicity, and add two more patients at this dose. Then we move onto the second stage with the standard 3+3 design. These three patients are treated

starting from the previous dose level. This procedure provides the opportunity to save time at the beginning of the trial and to avoid a large number of patients treated at a very low ineffective dose that is much lower than the MTD.

1.3 Parametric methods

1.3.1 Continual reassessment method

Rule-based methods to dose-finding only follow some pre-specified rules. We collect information based on the current dose, and we do not collect information on other doses. To overcome the problems of nonparametric methods, we use model-based methods called parametric methods in Phase I trials. First, we introduce the continual reassessment method (CRM), which is a commonly used method in a dose-finding design.

O'Quigley et al. (1990) propose the CRM which connects the true toxicity probabilities $\pi_1, \pi_2, \cdots, \pi_J$ with pre-specified toxicity probabilities $p_1, p_2, \cdots, p_J$ at each dose by a parametric model with an unknown parameter α. So the dose-finding decision making is based on the model that we define in the trial study. In general, we assume that the true toxicity probability π_i is a function of the dose level d_i. Moreover, the pre-specified toxicity probability p_i is monotonically nondecreasing in the dose level d_i. That is, $p_1 < p_2 < \cdots < p_J$. Let θ denote the target toxicity probability. The CRM introduced in O'Quigley and Shen (1996)

assumes the following model,

$$P(\text{toxicity at dose level } i) = \pi_i = p_i^{exp(\alpha)},$$

where α is the unknown parameter in the model.

There are some modifications of the CRM model. For example, a logistic model or a hyperbolic tangent model is widely used in the literature. The CRM initially introduced in O'Quigley et al. (1990) has been extended and improved in many directions, often in terms of different dose-toxicity models. For example, O'Quigley and Shen (1996) introduces a one-parameter logistic function

$$\pi_i(\alpha) = \frac{exp(\beta + \alpha d_i)}{1 + exp(\beta + \alpha d_i)},$$

where β is a fixed constant and d_i is the standardized dose level i. In particular, Yin (2012) applies a fixed constant β of -3. A hyperbolic tangent function,

$$\pi_i(\alpha) = \left\{\frac{\tanh(d_i) + 1}{2}\right\}^{\alpha} = \left\{\frac{(e^{2d_i} - 1)/(e^{2d_i} + 1) + 1}{2}\right\}^{\alpha},$$

was also introduced in O'Quigley et al. (1990).

The book by Cheung (2011) is a comprehensive study on CRM. The important issue of escalation with overdose control (EWOC) is considered by many authors, including Zacks et al. (1998), Babb et al. (1998), Tighiouart et al. (2005), Tighiouart and Rogatko (2010) and Chen et al. (2012). Wheeler et al. (2017) consider toxicity dependent feasibility bounds for EWOC. O'Quigley and Shen (1996) consider the likelihood approach for CRM and Yin and Yuan (2009b) apply

multiple CRMs to model the toxicity probability. Pan and Yuan (2017) introduce a default method to specify the skeleton for the Bayesian averaging CRM and Lee and Cheung (2009) consider model calibration. The potential problem of model misspecification for CRM is investigated in Shen and O'Quigley (1996). The stopping rule for CRM is studied in O'Quigley and Reiner (1998), and the issue of early termination of CRM is considered in O'Quigley (2002). Leung and Wang (2002) apply the decision theory to CRM.

A tutorial on CRM is given by Garrett-Mayer (2006). Sverdlov et al. (2014) provide an excellent survey on phase I clinical trial designs, and Ratain et al. (1993) describe many important issues and difficulties on using phase I designs. Moreover, Rosenberger and Haines (2002) and Potter (2006) are excellent reviews of various designs of phase I clinical trials. Finally, O'Quigley (1992) considers the problem of statistical inference of the CRM design and Neuenschwander et al. (2008) examine some critical aspects of the Bayesian approach to the design of phase I clinical trials.

The performance of CRM is evaluated in different settings. For example, Iasonos et al. (2008) compare the performance of CRM with the nonparametric 3+3 design. Most recently, Zhou et al. (2018) introduce performance measures of accuracy, safety and reliability. Cheung (2014) introduces the benchmark measure based on sample mean toxicity probabilities.

We can use frequentist or Bayesian approach to estimate the unknown pa-

rameter α and then MTD toxicity. In this Chapter, we only focus on the Bayesian method. Suppose that n_i patients are treated at dose level i, for $i = 1, 2, \cdots, J$, and y_i patients of the total of n_i patients experience toxicity. Assume that D is the observed data set. The likelihood function is

$$L(D|\alpha) \propto \prod_{i=1}^{J} \{p_i^{\exp(\alpha)}\}^{y_i} \{1 - p_i^{\exp(\alpha)}\}^{n_i - y_i}.$$

We assume that α follows a specified prior distribution, denoted by $f(\alpha)$. By the Bayes' theorem, the posterior mean toxicity probability at dose level i can be estimated as

$$\hat{\pi}_i = \int p_i^{\exp(\alpha)} \frac{L(D|\alpha) f(\alpha)}{\int L(D|\alpha) f(\alpha) d\alpha} d\alpha.$$

After treating all patients, we can obtain the posterior mean toxicity probability at each dose level, and find the dose level whose posterior mean toxicity probability is nearest to the target toxicity probability θ. Then this dose is the MTD recommended for the Phase II trial.

1.3.2 Bayesian model averaging CRM

In the CRM, we try to model the true toxicity probability. If our pre-specified toxicity probabilities are too far from the true toxicity probabilities, the estimates may not be precise, and the design may not perform well. To find the MTD in the Phase I trial, Yin and Yuan (2009b) apply multiple CRMs, each has a different prior, to model the true toxicity probability. Hoeting et al. (1999) propose a discrete prior probability to each CRM and assign each CRM model a weight

called a Bayeasian model averaging (BMA) procedure. In practice, we may assign a higher weight to the better fitted model and a lower weight to a poorer fitted model. So the estimates of toxicity probabilities approach closest to the best fitted model all over CRMs. In general the performance of BMA is better than a single CRM.

The BMA-CRM design uses multiple CRM models. Suppose M_k denotes the k^{th} CRM model related to a toxicity probability set, called a skeleton, $(p_{k1}, p_{k2}, \cdots, p_{kJ})$, where $k = 1, 2, \cdots, K$, and K is the total number of CRM models. Then the toxicity probability at d_i is

$$\pi_{ki}(\alpha_k) = p_{ki}^{\exp(\alpha_k)},$$

where $i = 1, 2, \cdots, J$, J is the total number of dose levels, and α_k is the unknown parameter related to CRM model M_k (Yin, 2012). Suppose that y_i patients who are treated over the total n_i patients experience toxicity. Assume that D is the observed data set. The likelihood function of CRM model M_k is

$$L(D|\alpha_k, M_k) \propto \prod_{i=1}^{J} \{p_{ki}^{\exp(\alpha)}\}^{y_i} \{1 - p_{ki}^{\exp(\alpha)}\}^{n_i - y_i}.$$

Then assume that the unknown parameter α follows a specified distribution, denoted by $f(\alpha_k|M_k)$, for the model M_k. The likelihood function of the model M_k is

$$L(D|M_k) = \int L(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k,$$

and the posterior model probability for M_k is

$$P(M_k|D) = \frac{L(D|M_k)P(M_k)}{\sum_{i=1}^{K} L(D|M_i)P(M_i)}.$$

Finally, the toxicity probability at dose level j is estimated by the Bayesian model averaging method as

$$\bar{\pi}_j = \sum_{k=1}^{K} \hat{\pi}_{kj} P(M_k|D), \quad j = 1, 2, \cdots, J.$$

Here,

$$\hat{\pi}_{kj} = \int p_{kj}^{\exp(\alpha_k)} \frac{L(D|\alpha_k, M_k) f(\alpha_k|M_k)}{\int L(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k} d\alpha_k$$

is the posterior mean of the toxicity probability at dose level j, under the assumption of model M_k.

So the Bayesian model averaging estimate of the toxicity probability is a weighted average of the posterior means, where $\hat{\pi}_{kj}$ is the weight of $P(M_k|D)$. After treating patients at dose level j, the decision of whether to escalate or de-escalate depends on the value of $\bar{\pi}_j$. An important issue is to develop an algorithm to find MTD that is closest to the prescribed target toxicity probability.

1.3.3 Escalation with overdose control

Babb et al. (1998) develop the escalation with overdose control (EWOC) design to protect patients from overdose. Assume $y_i = 1$ if the patient is toxic, and $y_i = 0$ otherwise. Then the toxicity probability of dose level i is the function F of dose

level. Define

$$P(\text{toxicity probability of dose } i) = P(y_i = 1|dose = x_i) = F(\alpha + \beta x_i),$$

where α and β are unknown parameters. Suppose there are n patients involved in the trial study, and the observed data set is $Y = \{y_1, \cdots, y_n\}$. Then the likelihood function is given by

$$L(Y|\alpha, \beta) = \prod_{i=1}^{n} \{F(\alpha + \beta x_i)\}^{y_i} \{1 - F(\alpha + \beta x_i)\}^{1-y_i}.$$

Suppose M and θ denote the MTD and the target toxicity probability, respectively. If x_0 is assumed to be the lowest dose level, then

$$\theta = P(y_i = 1|dose = M) = F(\alpha + \beta M),$$

and

$$\pi_0 = P(y_i = 1|dose = x_0) = F(\alpha + \beta x_0).$$

So we can calculate α and β as follows:

$$\beta = \frac{F^{-1}(\theta) - F^{-1}(\pi_0)}{M - x_0},$$
$$\alpha = F^{-1}(\pi_0) - \beta x_0.$$

We can assume that $F^{-1}(x)$ is exponential, logistic or hyperbolic tangent function. After assuming the prior distributions of M and π_0, we can get the joint posterior distribution of M and π_0. Integrating out π_0, the marginal distribution of M, denoted as $G(x|Y)$, can be used to find the next dose level. We define

$$G(x|Y) = P(M \leq x|Y).$$

In the EWOC design, the minimal posterior expected loss is used to determine the appropriate dose to allocate,

$$\int L_\gamma(x, M) dG(M|Y),$$

where γ is a pre-specified threshold, and $L_\gamma(x, M)$ is an asymmetric loss function (Yin, 2012).

1.3.4 CRM with combination drugs

Applying only one drug may not be effective in treating such disease as cancer. We need two or more drugs combined to treat patients. Yin (2012) extends one drug CRM model to combination drugs. Patients are treated sequentially, one at a time. Consider a combination of two drugs A and B. Drug A has K pre-specified toxicity probabilities $a_1 < a_2 < \cdots < a_K$ for K dose levels, and drug B has L pre-specified toxicity probabilities $b_1 < b_2 < \cdots < b_L$ for L dose levels. A new CRM design with the following new model of dose toxicity probability, which incorporates interaction between the two drugs, is introduced:

$$P(\text{toxicity at dose levels } (a_i, b_j)) = \pi(\alpha, \beta, \gamma) = \frac{exp(\eta + \alpha a_i + \beta b_j + \gamma a_i b_j)}{1 + exp(\eta + \alpha a_i + \beta b_j + \gamma a_i b_j)},$$

where η is taken as a constant, α, β and γ are unknown parameters, and $\gamma a_i b_j$ is the interaction term of two drugs.

For the purpose of simulation, assume that the parameters α, β and γ are random but follows some prior distributions. The Bayesian approach is

applied to estimate the prior or posterior mean toxicity probability at each combination of dose levels $(i, j), i = 1, \cdots, K; j = 1, \cdots, L$. Suppose patient $n, n = 1, 2, \cdots, N$, is treated at dose combination $(a^{(n)}, b^{(n)})$ and toxicity outcome y_n is observed, where $y_n = 1$ if the patient is toxic and $y_n = 0$ otherwise. Let $D_n = \{(a^{(1)}, b^{(1)}), y_1; (a^{(2)}, b^{(2)}), y_2; \cdots; (a^{(n)}, b^{(n)}), y_n\}$ be the observed information on dose combination and its corresponding toxicity of all previously treated patients, and denote

$$\pi_i = \pi(\alpha, \beta, \gamma) = \frac{exp(\eta + \alpha a^{(i)} + \beta b^{(i)} + \gamma a^{(i)} b^{(i)})}{1 + exp(\eta + \alpha a^{(i)} + \beta b^{(i)} + \gamma a^{(i)} b^{(i)}))}, \quad i = 1, 2, \cdots, n.$$

The likelihood function is

$$L(\alpha, \beta, \gamma | D_n) = \prod_{i=1}^{n} \{\pi_i\}^{y_i} \{1 - \pi_i\}^{1-y_i},$$

With our new model, this likelihood function becomes

$$\begin{aligned} L(\alpha, \beta, \gamma | D_n) &= \prod_{i=1}^{n} \left\{ \frac{exp(\eta + \alpha a^{(i)} + \beta b^{(i)} + \gamma a^{(i)} b^{(i)})}{1 + exp(\eta + \alpha a^{(i)} + \beta b^{(i)} + \gamma a^{(i)} b^{(i)}))} \right\}^{y_i} \\ &\quad \times \left\{ 1 - \frac{exp(\eta + \alpha a^{(i)} + \beta b^{(i)} + \gamma a^{(i)} b^{(i)})}{1 + exp(\eta + \alpha a^{(i)} + \beta b^{(i)} + \gamma a^{(i)} b^{(i)}))} \right\}^{1-y_i}. \end{aligned}$$

To ensure that the toxicity probability is increasing in the dose level of one drug when the dose level of the other drug is fixed, we assume that α, β, γ are all positive. Suppose that α follows the prior distribution $f(\alpha)$, β follows the prior distribution $g(\beta)$, and γ follows the prior distribution $h(\gamma)$. By the Bayes' Theorem, after treating n patients, the posterior mean toxicity probability at dose

level (a_i, b_j) is estimated to be

$$\begin{aligned}\hat{\pi}(i,j) &= \int\int\int \frac{exp(\eta+\alpha a_i+\beta b_j+\gamma a_i b_j)}{1+exp(\eta+\alpha a_i+\beta b_j+\gamma a_i b_j)} \\ &\quad \times \frac{L(\alpha,\beta,\gamma|D_n)f(\alpha)g(\beta)h(\gamma)}{\int\int\int L(\alpha,\beta,\gamma|D_n)f(\alpha)g(\beta)h(\gamma)d\alpha d\beta d\gamma} d\alpha d\beta d\gamma.\end{aligned}$$

The new challenge of finding dose combination level in two combination drugs is as follows. In the combination drug trial, the toxicity order of two drugs is only partially known. For example, there are two drugs, each has 3 dose levels, say 3×3 combined dose levels. Combination dose of $(3,2)$ and $(2,3)$ are more toxic than that of $(2,2)$, but we do not know between combination dose of $(3,2)$ and $(2,3)$ which one is more toxic. If we observe that the drug combination $(2,2)$ is nontoxic, should we move right to $(3,2)$ or move up to $(2,3)$? So determining the toxicity order of $(3,2)$ and $(2,3)$ is the important issue in dose movements.

1.4 Summary

In this chapter, we have reviewed basic ideas of some commonly used nonparametric and parametric designs of Phase I clinical trials. To summarize, the goal of Phase I clinical trial is to assess the toxicity of the new drug and identify the maximum tolerated dose for later phases. Details of the continual reassessment method in our new model are given in Chapter 2. The Bayesian model averaging model is introduced in Chapter 3. In Chapter 4, an unknown parameter α and a constant β in the basic model are all considered as constants, but the variance of the normal distribution $\Phi(\cdot)$ is unknown in the CRM model. We extend the one drug trial

to combination drugs CRM for more complicated treatment in Chapter 5. In Chapter 6, we introduce the statistical inference of the unknown parameter α in CRM. In the last chapter, we conclude and discuss potential future work.

Chapter 2

The CRM with one drug in Phase I trials

2.1 Introduction

This chapter is based on our paper Zhang et al. (2019) which has been accepted for publication.

In the current literature, three models are used for the CRM: the power model, logistic model and hyperbolic tangent model. See Chapter 1 for a summary.

In principle, any increasing function with the range of $(0, 1)$ may be used for the CRM. In this chapter, we introduce a new model with such a property. This new function is constructed using the cumulative distribution of the normal distribution, but extends the logistic model. It can be used to describe the dose toxicity probability relationship.

In this chapter, we also introduce an efficiency measure and criteria BEARS:

Benchmark, Efficacy, Accuracy, Reliability, Safety. We use BEARS to evaluate the performance of our new CRM design and existing models.

2.2 A new design of CRM for one drug

Suppose that a new drug is investigated for its toxicity at K dose levels $d_1 < d_2 < \cdots < d_K$, and $p_i, i = 1, 2, \cdots, K$, are some pre-speficied skeleton probabilities.

We introduce a new dose toxicity probability function by

$$\pi_i = \pi_i(\alpha, \beta) = \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} = 2 - 2(1 + \Phi(\beta + \alpha d_i))^{-1},$$

where Φ is the cumulative distribution function of the normal distribution $N(\mu, \sigma^2)$. The parameter β is positive and is assumed to be given, which determines the location of the dose toxicity probability curve. However, the parameter $\alpha > 0$ determines the shape of the dose toxicity probability curve. We consider the cases of an unknown α in Chapters 2 and 3, and a known α in Chapter 4.

My motivation for introducing our new model is as follows. Currently available models including the power function, logistic function and hyperbolic tangent function all represent increasing functions with the range of $(0, 1)$. Although we can change the value of α to have different shapes of the dose toxicity probability relationship based on the CRM model, these functions are restricted in particularly forms (i.e., power, logistic and hyperbolic tangent). I wished to expand these classes of CRM models.

Our new model extends the logistic model by replacing the exponential function with the cumulative distribution function of the normal distribution. The advantage of our new model over the other models (power, logistic and hyperbolic tangent) is that our new model is much more broad and allows more options to model an increasing function with the range of $(0, 1)$. The difference between the logistic model and our model is that the dose toxicity probability relationship of our model can be adjusted by not only parameters α and β but also μ and σ^2. In summary, both β and μ affect the location of the CRM model, and α and σ^2 affect the shape of the CRM model.

By choosing different values of μ and σ^2 of the cumulative distribution function of the normal distribution, we are able to model more different shapes of the increasing dose toxicity probability relationship than the three currently existing models in the literature. At the beginning of the trial, the dose level is very low and ineffective for patients. So the dose toxicity probability curve should be flat. As dose level goes up, we hope that the dose toxicity probability curve is steep since the toxicity probability increases more sharply in dose level. This means that the toxicity probability is sensitive when the dose level is escalated. We focus on the toxicity probabilities of all doses and subsequently determine the MTD. In the latter stages of the trial, we hope that the dose toxicity probability curve is flat because high dose levels can be fatal. A small increase in the dose level may lead to death, so overdose control of high dosages is of significant concern. Hence it is safer to patients if we model a flat dose toxicity probability curve at high

dose levels. As a result, there is a small increase in toxicity probability when the dose level is escalated.

Based on our model, we can model steeper or flatter dose toxicity probability curves more flexibly with different values of μ and σ^2, together with different values of parameters α and β. Graphs (a) and (b) in Figure 2.1 show different shapes of the dose toxicity probability curve for different values of σ^2 and different values of α. Graphs (c) and (d) in Figure 2.1 show different location of the dose toxicity probability curve for different values of μ and different values of β.

The increasing monotonicity of the dose-toxicity probability function is easily checked by its first derivative $\frac{d\pi_i}{dd_i} = \frac{2\alpha\phi(\beta+\alpha d_i)}{(1+\Phi(\beta+\alpha d_i))^2} > 0$, where ϕ is the probability density function of the normal distribution $N(\mu, \sigma^2)$.

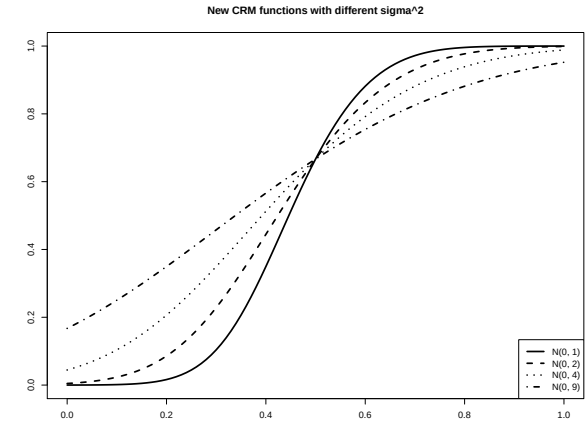
Furthermore, the probability of toxicity is increasing in the parameter α for every given dose checked by its first derivative is also important in the CRM formulation.

$$\begin{aligned}\frac{d\pi_i}{d\alpha} &= \frac{2\phi(\beta+\alpha d_i)d_i - 2\Phi(\beta+\alpha d_i)\phi(\beta+\alpha d_i)d_i}{(1+\Phi(\beta+\alpha d_i))^2} \\ &= \frac{2\phi(\beta+\alpha d_i)d_i(1-\Phi(\beta+\alpha d_i)d_i)}{(1+\Phi(\beta+\alpha d_i))^2} > 0,\end{aligned}$$

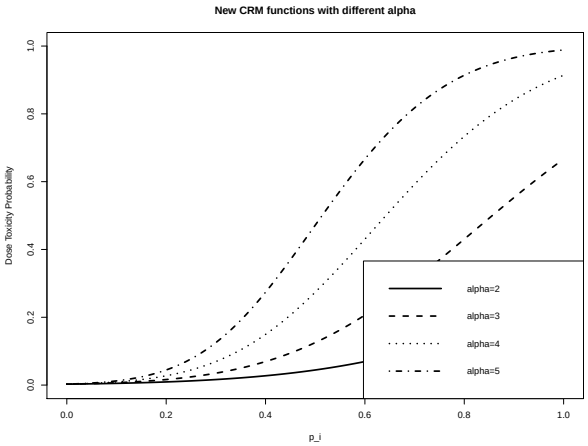
since the cumulative probability function $\Phi(\beta+\alpha d_i)$ is always less than 1.

The concavity of the dose toxicity probability function is determined by its second derivative

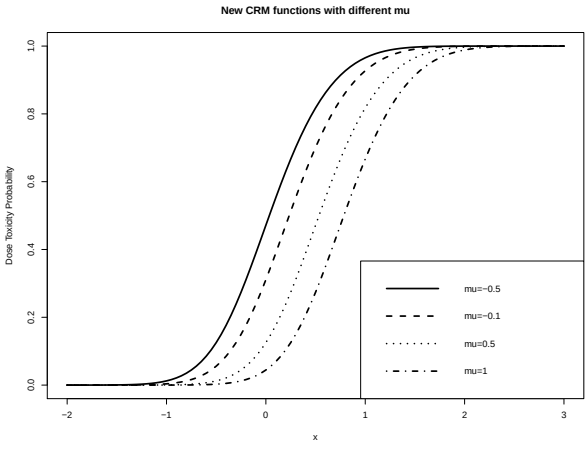
$$\frac{d^2\pi_i}{dd_i^2} = \frac{2\alpha^2\phi(\beta+\alpha d_i)'(1+\Phi(\beta+\alpha d_i)) - 4\alpha^2\phi(\beta+\alpha d_i)^2}{(1+\Phi(\beta+\alpha d_i))^3},$$



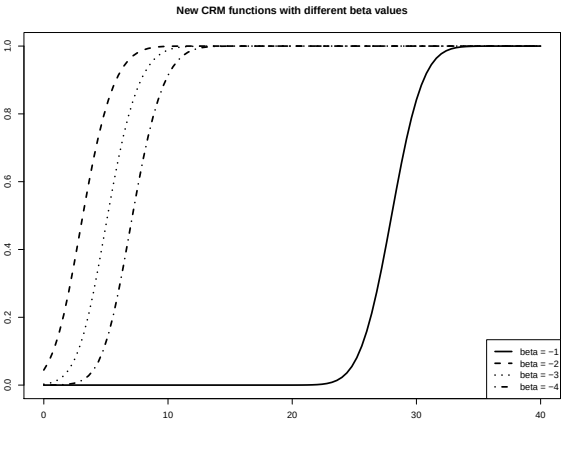
(a) New model for different parameter σ^2



(b) New model for different parameter α



(c) New model for different parameter μ



(d) New model for different parameter β

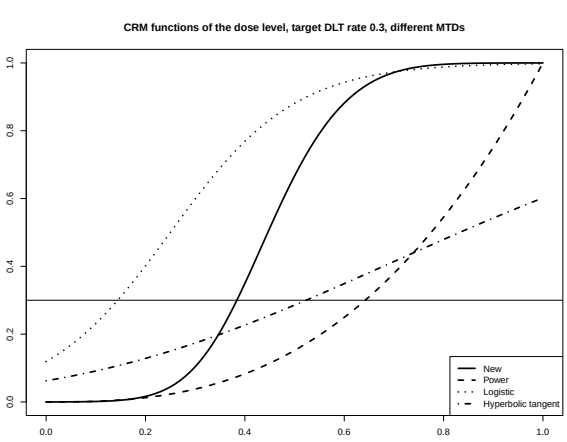
Figure 2.1: Various CRM functions based on our model

which is positive if $(1+\Phi)\phi' - 2\phi^2 \geq 0$. This is true if $\frac{\phi'}{\phi} \geq \frac{2\phi}{1+\Phi}$. So, if $(ln\phi)' \geq (2ln(1+\Phi))'$, the second derivative $\frac{d^2\pi_i}{dd_i^2}$ is positive. Because a logarithmic function is a one-to-one and monotonically increasing function, this is true if $\frac{\phi}{(1+\Phi)^2}$ is increasing in d_i. We have drawn the graphs of the dose toxicity probability function for different values of μ and σ^2, and observed that our new model depicts a variety of dose toxicity probability relations, just like the power, logistic and hyperbolic tangent functions.

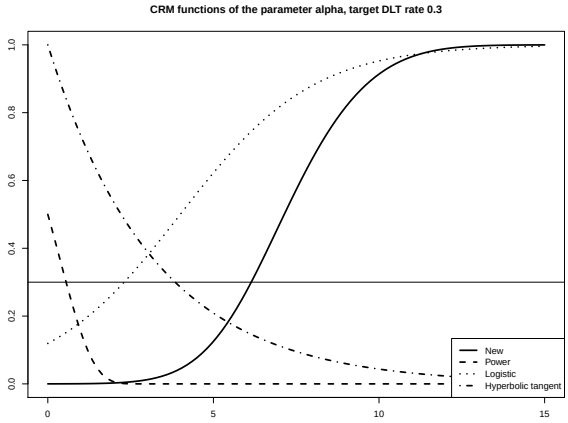
As a function of the dose level, the new dose toxicity probability function mimics that of the power, logistic and hyperbolic tangent functions, which is clearly seen in Figure 2.2, which shows the increasing monotonicity (in the dose level) of our new function and the power, logistic and hyperbolic tangent functions, and the (increasing or decreasing) monotonicity (in the parameter α) of these functions. The functions in Figure 2.2 (a) are $\frac{2\Phi(-4+8x)}{1+\Phi(-4+8x)}$ (new), $x^{exp(1)}$ (power), $\frac{exp(-2+8x)}{1+exp(-2+8x)}$ (logistic) and $\left(\frac{exp(x)}{exp(x)+exp(-x)}\right)^4$ (hyperbolic tangent). The functions in Figure 2.2 (b) are $\frac{2\Phi(-4+0.5x)}{1+\Phi(-4+0.5x)}$ (new), $0.5^{exp(x)}$ (power), $\frac{exp(-2+0.5x)}{1+exp(-2+0.5x)}$ (logistic) and $\left(\frac{exp(0.5)}{exp(0.5)+exp(-0.5)}\right)^x$ (hyperbolic tangent).

Dose-finding Algorithm

The CRM starts by treating the first cohort of patients at the safest dose. The cohort size can be three or one, depending on the design. Dose escalation or deescalation is determined by observed data and the posterior mean toxicity probabilities at all doses. A non-toxic response usually results in the same dose



(a) Functions of the dose level



(b) Functions of the parameter alpha, dose level 0.5

Figure 2.2: Various CRM functions: monotonicity

	True Toxicity Probabilities						Skeleton Probabilities						
Scenario	1	2	3	4	5	6	1	2	3	4	5	6	MTD
1	**0.280**	0.380	0.480	0.580	0.690	0.780	0.300	0.422	0.540	0.643	0.729	0.797	1
2	0.180	**0.280**	0.380	0.480	0.580	0.680	0.186	0.300	0.422	0.540	0.643	0.729	2
3	0.160	0.220	**0.300**	0.380	0.480	0.580	0.095	0.186	0.300	0.422	0.540	0.643	3
4	0.080	0.120	0.200	**0.300**	0.420	0.550	0.038	0.095	0.186	0.300	0.422	0.540	4
5	0.050	0.100	0.150	0.200	**0.280**	0.350	0.010	0.038	0.095	0.186	0.300	0.422	5
6	0.030	0.050	0.100	0.150	0.200	**0.250**	0.002	0.010	0.038	0.095	0.186	0.300	6

Table 2.1: Summary of simulation scenarios

level or escalation, while a toxic response usually results in the same dose level or deescalation. At the boundaries, dose escalation (at d_K) or deescalation (at d_1) does not occur. The size of escalation or deescalation is determined by the dose level whose posterior mean toxicity probability is nearest the target DLT rate. If skipping is not allowed, escalation or deescalation occurs at most one step away from the current dose.

When the m^{th} patient is to be assigned a dose level, Each increasing or decreasing dose is determined by previous observations and obtained by the posterior mean toxicity probabilities at all doses. Let $D = \{(d^{(1)}, y_1), \cdots, (d^{(m-1)}, y_{m-1})\}$ be the observed information when the m^{th} patient is to be treated, where $d^{(i)}$ is the dose applied to patient $i = 1, \cdots, m-1$ and $y_i = 1$ if toxicity is observed and 0 otherwise. The likelihood function becomes

$$L(D|\alpha) \propto \prod_{i=1}^{m-1} \left\{ \frac{2\Phi(\beta + \alpha d^{(i)})}{1 + \Phi(\beta + \alpha d^{(i)})} \right\}^{y_i} \left\{ 1 - \frac{2\Phi(\beta + \alpha d^{(i)})}{1 + \Phi(\beta + \alpha d^{(i)})} \right\}^{1-y_i}.$$

We assume the positive α follows a prior distribution $f(\alpha)$. By the Bayes' Theorem, the posterior mean toxicity probability $\hat{\pi}_i$ at dose level d_i is estimated to be

$$\hat{\pi}_i = \int \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} \frac{L(D|\alpha) f(\alpha)}{\int L(D|\alpha) f(\alpha) d\alpha} d\alpha.$$

After each cohort of patients is treated, we collect all toxicity data and calculate the posterior mean toxicity probabilities at all dose levels, say $\hat{\pi}_1, \hat{\pi}_2, \cdots, \hat{\pi}_K$. The dose whose posterior mean toxicity probability is closest to the target θ but still safe is applied to the next cohort of patients. The trial terminates when the

toxicity probabilities converge and the MTD is determined.

The dose finding procedure by CRM follows the following rules (Yin, 2012):

(1) We treat the first cohort of patients at the starting dose or the lowest dose.

(2) Let the current dose level be i^{cur}, and denote the target toxicity probability as θ. We can calculate the posterior means of the toxicity probabilities for all the observations, that is, $\hat{\pi}_1, \hat{\pi}_2, \cdots, \hat{\pi}_K$. We obtain the dose level i^* whose toxicity probability closest to θ,

$$i^* = argmin|\hat{\pi}_i - \theta|,$$

If $i^{cur} > i^*$, de-escalate to the next lower level, and if $i^{cur} < i^*$, escalate to the next higher level, otherwise, keep it at the same dose level.

(3) We determine the dose with the toxicity probability closest to θ as the MTD when the maximum sample size is collected.

2.3 Simulation results and discussion

We compare the performance of the following CRM designs: (1) the new models $I(a) = \frac{2\Phi(-1+\alpha p_i)}{1+\Phi(-1+\alpha p_i)}$, $I(b) = \frac{2\Phi(-1.5+\alpha p_i)}{1+\Phi(-1.5+\alpha p_i)}$ and $I(c) = \frac{2\Phi(-4+\alpha p_i)}{-4+\Phi(-1+\alpha p_i)}$ (representing good fit in the simulation scenarios); (2) the power model $II = p_i^{exp(\alpha)}$; (3) the logistic models $III(a) = \frac{exp(-1+\alpha p_i)}{1+exp(-1\alpha p_i)}$, $III(b) = \frac{exp(-2+\alpha p_i)}{1+exp(-2\alpha p_i)}$ and $III(c) = \frac{exp(-4+\alpha p_i)}{1+exp(-4\alpha p_i)}$ (representing good fit in the simulation scenarios); and (4) the hyperbolic tangent model $IV = \left(\frac{exp(p_i)}{exp(p_i)+exp(-p_i)}\right)^{\alpha}$.

The simulation setting is as follows: (1) the sample size is $n = 30$; (2) the target DLT rate is $\theta = 0.3$; (3) there are 6 dose levels and simulation scenarios are given in Table 2.1, together with their skeletons from the getprior() function; (4) the prior distribution for α is the $gamma(x, 0.5, 0.5)$ distribution in R; (5) each trial is simulated for 2,000 runs; and (6) the seed for the k^{th} simulation run, where $k = 1, 2, \cdots, 2000$, is set to be $2018 + k$ to ensure repeatability and consistency for all models.

The $gamma(x, 0.5, 0.5)$ density function is symmetric and suitable for our simulation. However, I have also run simulations with other prior distributions including the normal and expontial distributions. Results for the gamma distribution are the best and reoprted here.

Comparison of simulation results is based on criteria BEARS: Benchmark, Efficiency, Accuracy, Reliability, Safety. The benchmark is proposed in Cheung (2014), operating measures A1, A2, R1, R2, S1 and S2 are introduced in Zhou et al. (2018), and all other measures are new in this paper. The endpoint A1 is called the percentage of correct selection (PCS) in Zhou et al. (2018).

Scenario		Skipping is not permitted						Skipping is allowed					
(MTD)	Model	1	2	3	4	5	6	1	2	3	4	5	6
1	I(a)	**0.934**	0.064	0.003	0	0	0	**0.934**	0.064	0.003	0	0	0
(1)	II	0.538	0.361	0.078	0.020	0.003	0	0.521	0.366	0.091	0.018	0.004	0.001
	III(a)	0.821	0.152	0.024	0.004	0	0	0.821	0.152	0.025	0.003	0	0
	IV	0.751	0.196	0.047	0.007	0	0	0.752	0.200	0.043	0.007	0	0
2	I(b)	0.150	0.379	0.288	0.119	0.058	0.007	0.169	0.422	0.277	0.095	0.029	0.009
(2)	II	0.001	**0.556**	0.339	0.075	0.026	0.005	0	**0.498**	0.376	0.103	0.013	0.010
	III(b)	0.069	0.353	0.313	0.166	0.094	0.007	0.094	0.403	0.322	0.142	0.031	0.010
	IV	0.318	0.401	0.207	0.066	0.008	0.001	0.326	0.396	0.209	0.060	0.009	0.001
3	I(b)	0.020	0.179	0.347	0.291	0.143	0.022	0.021	0.179	0.393	0.270	0.091	0.047
(3)	II	0.001	0	**0.607**	0.319	0.059	0.015	0	0	**0.552**	0.324	0.083	0.041
	III(b)	0.002	0.091	0.336	0.334	0.196	0.042	0.005	0.115	0.377	0.306	0.152	0.047
	IV	0.193	0.290	0.305	0.158	0.052	0.004	0.202	0.268	0.304	0.174	0.048	0.005
4	I(c)	0.020	0	0.065	0.487	0.168	0.261	0.020	0	0.065	0.487	0.168	0.261
(4)	II	0.001	0	0	**0.707**	0.266	0.027	0.001	0	0	**0.630**	0.263	0.107
	III(c)	0.004	0	0.065	0.407	0.406	0.119	0.001	0	0.050	0.413	0.408	0.130
	IV	0.048	0.093	0.245	0.396	0.193	0.026	0.046	0.107	0.220	0.404	0.201	0.023
5	I(c)	0.001	0	0.008	0.168	0.467	0.358	0.001	0	0	0.034	**0.543**	0.433
(5)	II	0.001	0	0	0	**0.598**	0.402	0.001	0	0	0	0.481	0.519
	III(a)	0.002	0.001	0.171	0.401	0.314	0.113	0	0	0.087	0.428	0.345	0.141
	IV	0.023	0.037	0.090	0.208	0.325	0.318	0.025	0.019	0.114	0.177	0.350	0.317
6	I(c)	0.001	0	0	0.008	0.174	0.819	0.001	0	0	0	0.107	0.893
(6)	II	0.001	0	0	0	0	**0.999**	0.001	0	0	0	0	**1**
	III(c)	0.001	0	0	0	0.065	0.935	0.002	0	0	0	0.053	0.946
	IV	0.021	0.011	0.036	0.068	0.166	0.7	0.020	0.001	0.019	0.096	0.178	0.687

Table 2.2: Simulated values of Benchmark. Bold values are highest in each group.

Scenario		Efficiency		Accuracy			Reliability		Safety	
(MTD)	Model	E1	E2 (s.d.)	A1	A2 (s.d.)	A3	R1	R2	S1	S2 (s.d.)
1	I(a)	0	NAN (NAN)	**0.897**	**0.971** (0.082)	**0.997**	**0.001**	**0**	**0.104**	**0.006** (0.034)
(1)	II	0	NAN (NAN)	0.651	0.470 (0.328)	0.480	0.427	0.270	0.350	0.106 (0.139)
	III(a)	0	NAN (NAN)	0.755	0.886 (0.174)	0.939	0.006	0	0.245	0.023 (0.065)
	IV	0	NAN (NAN)	0.743	0.786 (0.262)	0.811	0.054	0.029	0.257	0.043 (0.095)
2	I(b)	0.158	0.172 (0.231)	0.493	0.362 (0.223)	0.304	0.073	**0.243**	0.350	0.117 (0.107)
(2)	II	**0**	**0.033** (0)	**0.66**	**0.475** (0.334)	**0.494**	0.396	0.280	0.341	0.123 (0.159)
	III(b)	0.073	0.093 (0.164)	0.493	0.327 (0.251)	0.299	0.138	0.360	0.435	0.145 (0.118)
	IV	0.267	0.457 (0.286)	0.437	0.291 (0.179)	0.118	**0.072**	0.257	**0.297**	**0.063** (0.111)
3	I(b)	0.210	0.119 (0.178)	0.456	0.317 (0.214)	0.224	0.074	0.305	0.335	0.149 (0.134)
(3)	II	**0**	**0.033** (0)	**0.701**	**0.521** (0.339)	**0.549**	0.331	**0.242**	0.299	0.137 (0.184)
	III(b)	0.122	0.071 (0.129)	0.448	0.294 (0.236)	0.239	0.124	0.390	0.431	0.188 (0.147)
	IV	0.386	0.275 (0.210)	0.367	0.245 (0.177)	0.079	**0.063**	0.363	**0.247**	**0.069** (0.122)
4	I(c)	0.067	**0.028** (0.107)	0.595	0.397 (0.349)	0.453	0.37	0.402	0.339	0.26 (0.286)
(4)	II	**0**	0.033 (0)	**0.792**	**0.561** (0.299)	**0.611**	0.256	**0.157**	**0.209**	0.170 (0.199)
	III(c)	0.074	0.047 (0.084)	0.475	0.234 (0.221)	0.140	0.464	0.501	0.452	0.313 (0.23)
	IV	0.312	0.158 (0.125)	0.451	0.309 (0.171)	0.130	**0.059**	0.215	0.238	**0.109** (0.145)
5	I(c)	0.176	0.069 (0.148)	0.594	0.371 (0.295)	0.394	0.320	0.386	**0.231**	0.355 (0.311)
(5)	II	**0**	**0.033** (0)	**0.644**	**0.483** (0.333)	**0.514**	0.392	**0.288**	0.357	0.384 (0.333)
	III(a)	0.385	0.170 (0.139)	0.380	0.244 (0.182)	0.088	**0.011**	0.340	0.236	**0.076** (0.132)
	IV	0.249	0.111 (0.100)	0.368	0.254 (0.174)	0.083	0.282	0.347	0.384	0.301 (0.290)
6	I(c)	0.177	0.060 (0.124)	0.823	0.698 (0.278)	0.821	0	0.131	0	NAN (NAN)
(6)	II	**0**	**0.033** (0)	**1**	**0.833** (0)	**1**	0	**0**	0	NAN (NAN)
	III(c)	0.067	0.041 (0.058)	0.934	0.798 (0.130)	0.941	0	0.007	0	NAN (NAN)
	IV	0.173	0.080 (0.079)	0.827	0.598 (0.245)	0.707	0	0.095	0	NAN (NAN)

Table 2.3: Simulated values of EARS, not allowing skipping. Bold values are the best in each group.

Scenario		Efficiency		Accuracy			Reliability		Safety	
(MTD)	Model	E1	E2 (s.d.)	A1	A2 (s.d.)	A3	R1	R2	S1	S2 (s.d.)
1	I(a)	0	NAN (NAN)	**0.897**	**0.971** (0.082)	**0.997**	**0.001**	**0**	**0.104**	**0.006** (0.034)
(1)	II	0	NAN (NAN)	0.634	0.454 (0.328)	0.457	0.418	0.290	0.366	0.109 (0.138)
	III(a)	0	NAN (NAN)	0.755	0.886 (0.174)	0.939	0.006	0	0.245	0.023 (0.065)
	IV	0	NAN (NAN)	0.742	0.785 (0.263)	0.811	0.057	0.027	0.259	0.043 (0.096)
2	I(b)	0.162	0.187 (0.245)	0.506	0.372 (0.231)	0.318	0.085	**0.230**	0.332	0.110 (0.110)
(2)	II	**0**	**0.033** (0)	**0.633**	**0.426** (0.335)	**0.435**	0.413	0.317	0.367	0.135 (0.156)
	III(b)	0.095	0.111 (0.185)	0.492	0.334 (0.262)	0.297	0.160	0.331	0.414	0.139 (0.123)
	IV	0.264	0.459 (0.285)	0.433	0.287 (0.177)	0.107	**0.076**	0.265	**0.304**	**0.064** (0.111)
3	I(b)	0.225	0.113 (0.193)	0.456	0.335 (0.233)	0.275	0.108	**0.287**	0.320	0.146 (0.143)
(3)	II	**0**	**0.017** (0)	**0.672**	**0.457** (0.345)	**0.481**	0.381	0.295	0.329	0.170 (0.184)
	III(b)	0.125	0.065 (0.15)	0.467	0.303 (0.259)	0.261	0.179	0.391	0.408	0.189 (0.158)
	IV	0.393	0.266 (0.210)	0.355	0.250 (0.181)	0.083	**0.061**	0.363	**0.252**	**0.073** (0.125)
4	I(c)	0.067	0.028 (0.107)	0.595	0.397 (0.349)	0.453	0.370	0.402	0.339	0.260 (0.286)
(4)	II	**0**	**0.011** (0)	**0.785**	**0.526** (0.304)	**0.569**	0.298	**0.173**	**0.216**	0.220 (0.202)
	III(c)	0.056	0.022 (0.077)	0.526	0.238 (0.252)	0.171	0.499	0.489	0.419	0.348 (0.255)
	IV	0.314	0.151 (0.122)	0.464	0.322 (0.17)	0.136	**0.063**	0.193	0.223	**0.113** (0.145)
5	I(c)	0.035	0.016 (0.072)	**0.700**	**0.425** (0.328)	**0.473**	0.447	0.326	0.266	0.513 (0.336)
(5)	II	**0**	**0.008** (0)	0.577	0.377 (0.336)	0.382	0.569	0.393	0.423	0.590 (0.336)
	III(a)	0.368	0.165 (0.145)	0.401	0.251 (0.186)	0.096	**0.026**	0.331	**0.232**	**0.088** (0.149)
	IV	0.252	0.103 (0.099)	0.368	0.28 (0.175)	0.104	0.270	**0.300**	0.381	0.309 (0.294)
6	I(c)	0.110	0.025 (0.111)	0.891	0.876 (0.248)	0.894	0	0.067	0	NAN (NAN)
(6)	II	**0**	**0.007** (0)	**1**	**0.967** (0)	**1**	0	**0**	0	NAN (NAN)
	III(c)	0.058	0.014 (0.062)	0.942	0.93 (0.139)	0.959	0	0.003	0	NAN (NAN)
	IV	0.192	0.074 (0.084)	0.809	0.630 (0.269)	0.697	0	0.088	0	NAN (NAN)

Table 2.4: Simulated values of EARS, allowing skipping. Bold values are the best in each group.

Scenario		Skipping is not permitted						Skipping is allowed					
(MTD)	Model	1	2	3	4	5	6	1	2	3	4	5	6
1	I(a)	**0.897**	0.096	0.005	0.003	0.001	0	**0.897**	0.096	0.005	0.003	0.001	0
(1)	II	0.651	0.316	0.032	0.002	0	0	0.634	0.333	0.031	0.003	0	0
	III(a)	0.755	0.214	0.025	0.005	0.001	0	0.755	0.214	0.025	0.006	0.001	0
	IV	0.743	0.211	0.044	0.003	0	0	0.742	0.210	0.047	0.002	0	0
2	I(b)	0.158	0.493	0.280	0.057	0.01	0	0.162	0.506	0.262	0.056	0.011	0.005
(2)	II	0	**0.660**	0.303	0.034	0.004	0	0	**0.633**	0.332	0.034	0.002	0
	III(b)	0.073	0.493	0.345	0.074	0.013	0.004	0.095	0.492	0.336	0.064	0.012	0.002
	IV	0.267	0.437	0.240	0.055	0.003	0	0.264	0.433	0.246	0.056	0.003	0
3	I(b)	0.013	0.197	0.456	0.248	0.062	0.026	0.019	0.207	0.456	0.248	0.051	0.022
(3)	II	0	0	**0.701**	0.267	0.030	0.002	0	0	**0.672**	0.293	0.034	0.002
	III(b)	0	0.122	0.448	0.329	0.080	0.023	0.001	0.124	0.467	0.320	0.067	0.022
	IV	0.128	0.258	0.367	0.203	0.040	0.005	0.128	0.265	0.355	0.209	0.039	0.005
4	I(c)	0	0	0.067	0.595	0.034	0.305	0	0	0.067	0.595	0.034	0.305
(4)	II	0	0	0	**0.792**	0.203	0.006	0	0	0	**0.785**	0.207	0.009
	III(c)	0	0	0.074	0.475	0.346	0.106	0	0	0.056	0.526	0.302	0.117
	IV	0.007	0.041	0.265	0.451	0.213	0.025	0.006	0.045	0.263	0.464	0.204	0.019
5	I(c)	0	0	0.004	0.172	0.594	0.231	0	0	0	0.035	**0.700**	0.266
(5)	II	0	0	0	0	**0.644**	0.357	0	0	0	0	0.577	0.423
	III(a)	0	0	0.055	0.330	0.380	0.236	0	0	0.054	0.314	0.401	0.232
	IV	0.001	0.004	0.047	0.198	0.368	0.384	0.001	0.005	0.047	0.199	0.368	0.381
6	I(c)	0	0	0	0.004	0.173	0.823	0	0	0	0	0.110	0.891
(6)	II	0	0	0	0	0	**1**	0	0	0	0	0	**1**
	III(c)	0	0	0	0	0.067	0.934	0	0	0	0	0.058	0.942
	IV	0	0.002	0.004	0.035	0.133	0.827	0	0.001	0.006	0.035	0.151	0.809

Table 2.5: Simulated percentages of MTD selection. Bold values are the highest in each group.

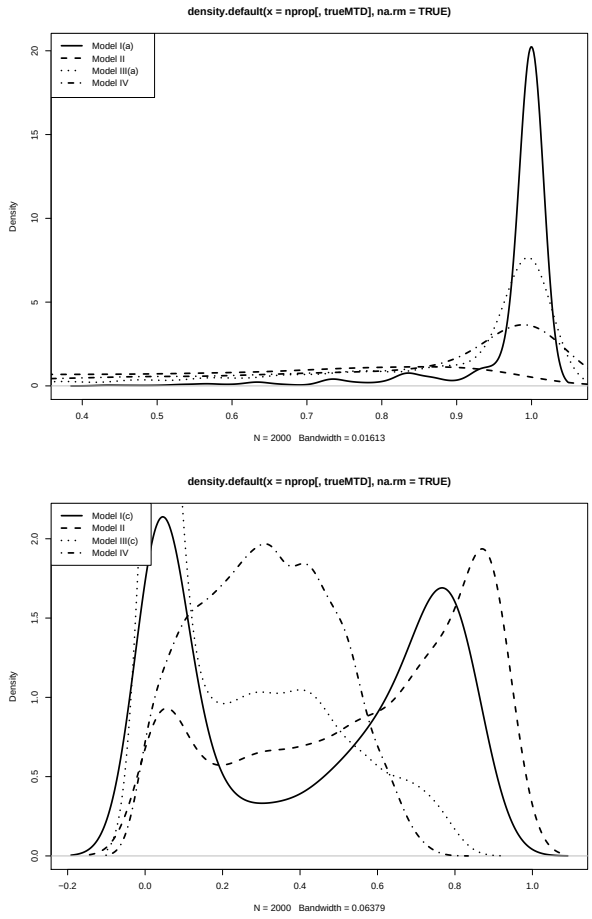


Figure 2.3: Comparison of density estimation of MTD allocation, scenarios 1 and 4, no skipping.

Benchmark means using sample toxicity probabilities to choose the MTD.

- **Performance measure:** The dose with a sample mean toxicity probability nearest to θ is chosen as the MTD. The higher the proportion of correctly selecting the MTD, the better the design.
- **Results:** From Table 2.2, the new model performs the best when the MTD is the lowest dose. In most other cases, the new model is the second best. Overall the new design offers good benchmark.

Efficiency checks the avoidance of assigning patients to ineffective doses.

- **Two performance measures:** E1 reports the proportion of simulation runs that assign patients to any dose lower than the MTD. The lower the proportion E1, the better the design. E2 gives the mean and standard deviation of the percentage of patients assigned to any dose lower than the MTD. The lower the mean, the better the design.
- **Results:** From Tables 2.3 and 2.4, the new model gives the lowest or second lowest values in many cases. Overall the new design is efficient.

Accuracy depicts the assignment of patients to the MTD.

- **Three performance measures:** A1 measures the proportion of simulation runs that correctly select the MTD. The higher the proportion, the better the design. A2 reports the mean and standard deviation of the percentage

of patients assigned to the MTD. The larger the mean, the better the design. A3 calculates the proportion of simulation runs with more than 50% of patients assigned to the MTD. The higher the proportion, the better the design.

- **Results:** From Tables 2.3 and 2.4, the new model gives either the highest or second highest values in most cases. Overall the new design is accurate.

Reliability concerns the risk of severe overdosing.

- **Two performance measures:** R1 measures the proportion of simulation runs that allocate more than 50% of patients to any dose higher than the MTD. The lower the proportion, the better the design. R2 calculates the proportion of simulation runs with less than one-sixth of patients at the MTD. The lower the proportion, the better the design.

- **Results:** From Tables 2.3 and 2.4, the new model gives either the lowest or second lowest values in many cases. Overall the new design is reliable.

Safety refers to the protection from overdosing.

- **Two performance measures:** S1 measures the proportion of simulation runs that result in any dose higher than the MTD. The lower its value, the better the design. S2 reports the mean and standard deviation of the percentage of patients assigned to any dose higher than the MTD. The lower the mean, the better the design.

- **Results:** From Tables 2.3 and 2.4, the new model either gives the lowest values in some cases or is reasonably small in other cases. Overall the new design is safe.

Table 2.5 summarizes the distribution of the MTD whose last updated posterior mean toxicity probability is nearest to the target DLT rate θ. In most cases the new design gives either the highest or second highest proportion of correct MTD identification.

Finally, Figure 2.3 compares the density estimation of the MTD selection out of 2,000 simulation runs. The new design offers high frequencies of large MTD allocations.

2.4 Application

We now illustrate the performance of our new model by a real application. This example of a real clinical trial is taken from the publication Yin and Yuan (2009b). This is a Phase I clinical trial for prostate cancer and was conducted at M.D. Anderson Center. This clinical trial investigated six doses of 20, 25, 30, 35, 40 and 45 mg/m^2 weekly for 4 weeks. The target toxicity probability was 40%. The true toxicity probabilities of six dose levels were 0.01, 0.05, 0.20, 0.40, 0.60, 0.85. Three skeletons of six dose levels were listed in Table 2.6. In Table 2.7, only skeleton 2 correctly selected the MTD as dose 4. 80% of total patients were assigned at the MTD in our model which is the best. In scenario 1, 75.5% of total patients

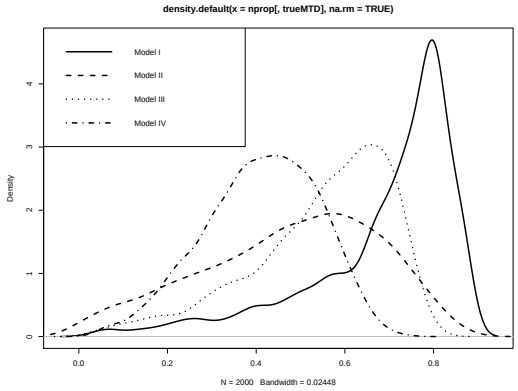
Scenario	MTD	True Toxicity Probabilities						Skeleton Probabilities					
		1	2	3	4	5	6	1	2	3	4	5	6
1	4	0.010	0.05	0.200	0.400	0.600	0.850	0.300	0.400	0.500	0.600	0.700	0.800
2	4	0.010	0.05	0.200	0.400	0.600	0.850	0.070	0.160	0.300	0.400	0.460	0.530
3	4	0.010	0.05	0.200	0.400	0.600	0.850	0.010	0.050	0.100	0.150	0.200	0.400

Table 2.6: True toxicity probabilities and skeletons for application

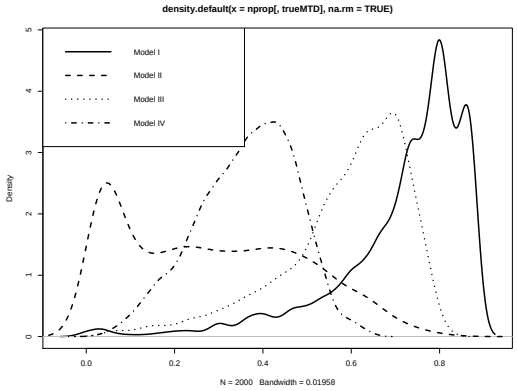
Scenario (MTD)	Model	Skipping is not permitted					
		1	2	3	4	5	6
1 (4)	I	0.000	0.002	0.072	0.755	0.166	0.007
	II	0.000	0.001	0.183	0.664	0.150	0.003
	III	0.000	0.002	0.082	0.698	0.210	0.009
	IV	0.001	0.029	0.186	0.597	0.182	0.006
2 (4)	I	0.000	0.005	0.043	**0.800**	0.151	0.003
	II	0.001	0.000	0.000	0.628	0.358	0.014
	III	0.000	0.000	0.048	0.736	0.209	0.008
	IV	0.002	0.029	0.186	0.576	0.203	0.006
3 (4)	I	0.034	0.002	0.106	0.129	0.700	0.031
	II	0.830	0.007	0.000	0.000	0.000	0.164
	III	0.003	0.000	0.055	0.234	0.680	0.029
	IV	0.001	0.038	0.219	0.518	0.222	0.004

Table 2.7: Simulated values of Benchmark.

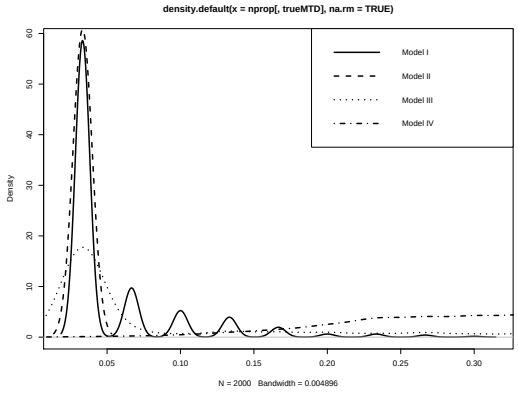
were assigned at the dose 4. But in scenario 1, dose 2 is the pre-specified MTD in our initial guess. This means that we under-guessed the effective dose level. In scenario 3, we assumed dose 6 is the MTD, but 70% of total patients were treated at dose 4 in our model. This means that we over-guessed the MTD. In Table 2.8, 96.7% of total patients were truly treated at the MTD in our model in scenario 2, which is the best. Finally, in Figure 2.4, graphs (a), (b) and (c) compare the density estimation of the MTD selection out of 2,000 simulation runs in three scenarios.



(a) scenario 1



(b) scenario 2



(c) scenario 3

Figure 2.4: Comparison of density estimation of MTD allocation, scenario 1 to 3, no skipping

Scenario (MTD)	Model	Skipping is not permitted					
		1	2	3	4	5	6
1 (4)	I	0.000	0.000	0.004	0.956	0.029	0.012
	II	0.000	0.000	0.241	0.622	0.138	0.000
	III	0.000	0.002	0.034	0.917	0.046	0.003
	IV	0.000	0.004	0.191	0.595	0.204	0.006
2 (4)	I	0.000	0.006	0.000	**0.967**	0.021	0.008
	II	0.000	0.000	0.000	0.671	0.329	0.001
	III	0.000	0.000	0.016	0.956	0.031	0.001
	IV	0.000	0.003	0.242	0.519	0.219	0.019
3 (4)	I	0.000	0.001	0.167	0.012	0.759	0.062
	II	0.000	0.000	0.000	0.000	0.000	1.000
	III	0.000	0.001	0.101	0.156	0.737	0.006
	IV	0.003	0.043	0.221	0.437	0.292	0.006

Table 2.8: Simulated values of Final MTD selection.

2.5 Conclusion

Phase I clinical trials are fundamentally important in drug development and their common goal is to determine the maximum tolerated dose while exposing fewer patients to toxic or ineffective doses. There are both nonparametric and parametric approaches to designing phase I clinical trials and each design has its own pros and cons. The performance of a parametric CRM design depends on the dose toxicity probability function.

We introduce a new dose toxicity probability function for CRM and evaluate its performance based on the template of BEARS (Benchmark, Efficiency, Accuracy, Reliability, Safety). The new CRM design behaves reasonably well overall and achieves the important goals of reliably identifying the maximum tolerated dose and safely minimizing risk.

Chapter 3

BMA-CRM

3.1 Introduction

Everyone's eyes may have a different view of a statistical model, and every statistician may choose a different skeleton. As we can see from the last chapter, the skeleton plays an important role for the statistical design of Phase I clinical trials. However, the pre-specified toxicity probabilities, called the skeleton, is subjective and so different statisticians may have different choices of skeleton. Since we may not know toxicity information of the new drug, how to choose a fitted model becomes an important issue in Phase I clinical trials. If the pre-specified toxicity probability deviates far from the true toxicity probability, the estimation of true toxicity probability may not be accurate, and the performance of the design may not be good. This may result in a wrong MTD. In practice, the true toxicity probabilities are unknown to statisticians, so there is no information to verify which skeleton is reasonable. To overcome this weakness and increase the

reliability of the trial design, Raftery et al. (1997) and Hoeting et al. (1999) assign each posterior probability a weight. Yin and Yuan (2009b) propose multiple CRM models, each one with a different skeleton. Pan and Yuan (2017) address the default method to specify skeletons for Bayesian model averaging (BMA). To address the issue of skeleton specification in this chapter, we assign a discrete prior probability to each CRM and estimate the toxicity and efficacy probabilities using the BMA model.

The implementation of CRM depends of the choice of a skeleton which is incorporated into the model. Lee and Cheung (2009) explore the issue of model calibration in CRM. To reduce the sensitivity of CRM on its skeleton, Yin and Yuan (2009b) introduce the Bayesian-model averaging CRM (BMA-CRM) by applying Bayesian model averaging to CRM, and Pan and Yuan (2017) introduce a default method to specify the skeletons for BMA-CRM. A thorough summary and performance comparison of CRM designs is provided in Zhou et al. (2018) and based on the performance metrics of MTD selection, patient allocation and overdose control.

3.2 Bayesian Model Averaging-CRM

Yin and Yuan (2009b) address the Bayesian model averaging method under the power model for the CRM. They introduce the BMA method to reduce dependence between model estimates of the toxicity probabilities for each dose level and assign

a weighted average toxicity probability among different models which can lead to a prediction better than a single CRM model.

Let M_n denote the n^{th} CRM toxicity probability model, where $n = 1, 2, \cdots, N-1, N$. Here, we have N models. Let $f(M_n)$ be the prior probability of model M_n and $f(M_n|D)$ be the posterior probability of the M_n given observed data set D. So the posterior probability of M_n is given by

$$f(M_n|D) = \frac{f(D|M_n)f(M_n)}{\sum_{i=1}^{N} f(D|M_i)f(M_i)}.$$

where $n = 1, 2, \cdots, N-1, N$.

The marginal likelihood function under M_n is

$$L(D|M_n) = \int L(D|\alpha_n, M_n)f(\alpha_n|M_n)d\alpha_n$$

where $f(\alpha_n|M_n)$ is the prior distribution of α_n under the M_n, and α_n is the unknown parameter in the model M_n.

Given observation D, the BMA estimate of the toxicity probability at dose level i is given by

$$\bar{\pi}_i = \sum_{n=1}^{N} \hat{\pi}_{ni} f(M_n|D),$$

where $i = 1, 2, \cdots, k$, and $\hat{\pi}_{ni}$, the posterior mean of the toxicity probability at dose level i under M_n, is defined as

$$\hat{\pi}_{ni} = \int \frac{2\Phi(\beta_n + \alpha_n p_i)}{1 + \Phi(\beta_n + \alpha_n p_i)} \frac{L(D|\alpha_n, M_n)f(\alpha_n|M_n)}{\int L(D|\alpha_n, M_n)f(\alpha_n|M_n)d\alpha_n} d\alpha_n.$$

where $(p_1, \cdots, p_k)$ is the skeleton. Therefore, $\bar{\pi}_i$ is a weighted posterior toxicity probability at dose level i for M_n. The BMA-CRM assigns a higher weight to a better fitted model and assigns a lower weight to a less fitted model. The escalation or de-escalation rule depends on not only the prior distribution but also $\bar{\pi}_i$. The BMA-CRM design becomes a problem of how to identify the MTD under the model fitting. This method can also be extended to combination drugs by introducing

$$\bar{\pi}_{ij} = \sum_{n=1}^{N} \hat{\pi}_{ij}^{(n)} f(M_n|D),$$

where $\bar{\pi}_{ij}$ is the posterior mean of the toxicity probability at combination dose (a_i, b_j) under M_n.

Dose-Finding Algorithm

Let θ be the target toxicity probability, say 30%. For the sake of safety to patients in the trial, we treat the first patient at the the lowest dose, and escalate or de-escalate only one dose level at one time. The dose-finding algorithm for our BMA-CRM design is as follows:

(1) We treat the first cohort of patients at the lowest dose level.

(2) Let the current dose level be i^{cur}. We calculate the posterior means of all toxicity probabilities under the BMA-CRM method , $\bar{\pi}_i$, $i = 1, 2, \cdots, K$. We decide the dose level i^* whose toxicity probability is closest to θ. That

is,

$$i^* = argmin|\bar{\pi}_i - \theta|,$$

If $i^{cur} > i^*$, we de-escalate to the next lower level. If $i^{cur} < i^*$, we escalate to the next higher level. Otherwise, the dose level does not change.

(3) We determine the dose with its toxicity probability closest to θ as the MTD, when the maximum sample size is collected.

3.3 Simulation Study

We compare the performance of individual CRMs with BMA-CRM under the Bayesian Model Averaging procedure. The simulation setting is as follows: (1) the sample size is $n = 30$; (2) the target DLT rate is $\theta = 0.3$; (3) there are 6 dose levels and 6 simulation scenarios, which are given in Table 3.1, together with 6 skeletons indicating 6 different CRM models, which are named $CRM1, CRM2, \cdots, CRM6$; (4) the prior distribution for α is the $gamma(x, 0.5, 0.5)$ distribution in R; (5) each trial is simulated for 2,000 runs.

In Table 3.1, the true toxicity probabilities are listed in the 2^{nd} column and the bold value in each row is the MTD in each scenario. The skeletons are listed in the 3^{rd} column. The dose toxicity probabilities increase evenly in dose at first two skeletons. For the third and fourth skeletons, toxicity probabilities increase slowly at the first three doses. For the last two skeletons, the toxicity probabilities are very low, even at the highest dose level. We have different characteristics in

the six skeletons so that we expect to obtain the true toxicity probabilities more effectively with BMA than with individual CRMs.

The last section shows that our new model performs better than the existing models in the literature according to "BEARS". In this section, comparison of simulation results is also based on this criteria: Benchmark, Efficacy, Accuracy, Reliability and Safety.

Benchmark means using sample mean toxicity probabilities to choose the MTD. In Table 3.2, we compare benchmark of dose selection probability at the MTD using posterior mean toxicity probabilities.

- **Performance measure:** The dose with a sample mean toxicity probability nearest to θ is chosen as the MTD. The higher the proportion of correct selection of the MTD, the better the design.

- **Results:** From Table 3.2, we see that the MTD is at dose level 1 in scenario 1. CRM1, CRM6 and BMA-CRM correctly identify the true MTD by giving the highest percentages of benchmark, and CRM1 and BMA-CRM correctly identify the true MTD by giving the highest percentages of final MTD selection to dose 1. In scenario 2, only CRM2 correctly identifies the true MTD of dose 2. CRM1, CRM3 and CRM4 instead identify dose 3 as the MTD. CRM5 and CRM6 identify dose 4 as the MTD. As a result, BMA-CRM, which is the average of CRM1 to CRM6, is heavily affected by the remaining CRMs and consequently misidentifies the true MTD. In

scenario 3, CRM1, CRM 3 and BMA correctly identify the true MTD at dose 3, and Benchmark and MTD selection percentages using BMA-CRM are the highest among the seven models. In scenario 4, CRM2, CRM4, CRM5 and BMA-CRM correctly identify the true MTD at the dose 4. CRM 6 is the worst MTD selection with 10.1% for Benchmark and 9.2% for the final MTD selection. In scenario 5, CRM 1, CRM3 and CRM5 correctly identify the true MTD at dose 5. The remaining CRMs misidentify the true MTD. In particular, CRM2 and CRM4 are the worst two models, which contribute more weight to BMA-CRM. Therefore, BMA-CRM dose not perform well. In scenario 6, all of the 7 models perform well, and correctly identify the true MTD at dose 6.

Efficiency checks the avoidance of assigning patients to ineffective doses.

- **Two performance measures:** E1 reports the proportion of simulation runs that assign patients to any dose lower than the MTD. The lesser the proportion E1, the better the design. E2 gives the mean and standard deviation of the percentage of patients assigned to any dose lower than the MTD. The lower the mean, the better the design.

- **Results:** From Tables 3.3, we see that BMA-CRM dose not give the lowest values of E1 and E2. However, these values are relatively low. Therefore, overall, the BMA-CRM design is efficient.

Accuracy depicts the assignment of patients to the MTD.

- **Three performance measures:** A1 measures the proportion of simulation runs that correctly select the MTD. The larger the proportion, the better the design. A2 reports the mean and standard deviation of the percentage of patients assigned to the MTD. The larger the mean, the better the design. A3 calculates the proportion of simulation runs with more than 50% of patients assigned to the MTD. The higher the proportion, the better the design.

- **Results:** From Tables 3.3, BMA-CRM gives reasonably high values of A1, A2 and A3 in all scenarios, and particularly in scenario 1, 3, 4 and 6. In scenario 2 and 5, BMA-CRM dose not seem perform very well. The reason may be the same as that for the Benchmark and final MTD selection percentage. Overall, BMA-CRM is accurate.

Reliability concerns the risk of severe overdosing.

- **Two performance measures:** R1 measures the proportion of simulation runs that allocate more than 50% of patients to any dose higher than the MTD. The lower the proportion, the better the design. R2 calculates the proportion of simulation runs with less than one-sixth of patients at the MTD. The lower the proportion, the better the design.

- **Results:** From Tables 3.3, BMA-CRM performs above average in all sce-

narios. Particularly BMA-CRM is the second best design in scenario 1, scenario 2 and scenario 3. Overall, the BMA-CRM design is reliable.

Safety refers to the protection from overdosing.

- **Two performance measures:** S1 measures the proportion of simulation runs that result in any dose higher than the MTD. The lesser its value, the better the design. S2 reports the mean and standard deviation of the percentage of patients assigned to any dose higher than the MTD. The lower the mean, the better the design.

- **Results:** From Tables 3.3, BMA-CRM performs the second best or third best with similar values except for scenario 5. In scenario 5, BMA-CRM seems performing not very well. The reason may be the same as that for the Benchmark and final MTD selection percentage. Overall, the BMA-CRM design is safe.

Putting all criteria "BEARS" together, the performance of the BMA-CRM design is overall stable. This is particularly valuable when we have no idea about which dose may be the true MTD. In such a case, the BMA-CRM design does not depend on this knowledge and is robust.

Finally, Figure 3.2 compares the density estimation of the MTD selection out of 2,000 simulation runs. The BMA-CRM design offers relatively high frequencies of MTD allocations.

	Toxicity Probabilities						Skeleton Probabilities						
Scenario	1	2	3	4	5	6	1	2	3	4	5	6	MTD
1	**0.280**	0.380	0.480	0.580	0.690	0.780	0.300	0.422	0.540	0.643	0.729	0.797	1
2	0.180	**0.280**	0.380	0.480	0.580	0.680	0.186	0.300	0.422	0.540	0.643	0.729	2
3	0.160	0.220	**0.300**	0.380	0.480	0.580	0.095	0.186	0.300	0.422	0.540	0.643	3
4	0.080	0.120	0.200	**0.300**	0.420	0.550	0.038	0.095	0.186	0.300	0.422	0.540	4
5	0.050	0.100	0.150	0.200	**0.280**	0.350	0.010	0.038	0.095	0.186	0.300	0.422	5
6	0.030	0.050	0.100	0.150	0.200	**0.250**	0.002	0.010	0.038	0.095	0.186	0.300	6

Table 3.1: Summary of simulation scenarios

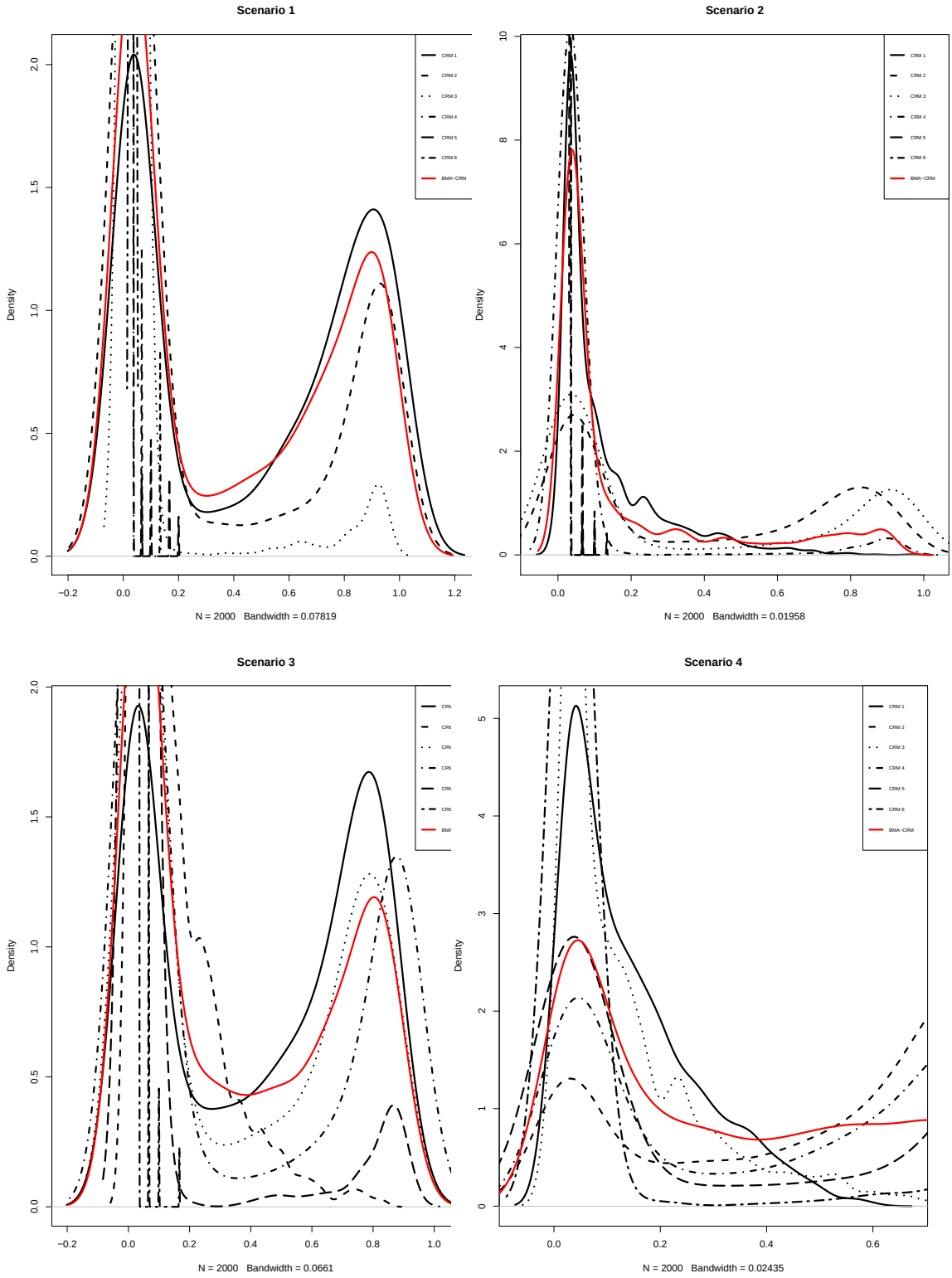
3.4 Conclusion

In practice, the performance of the CRM depends heavily on the choice of the skeleton. However, the choice of the skeleton is often subjective and particular choice is not necessarily the best. Therefore, the performance of the CRM may not be robust.

In this chapter, we take the approach of the BMA-CRM and consider all possible choice of the skeleton. This may avoid the possibility of the CRM depending on the single skeleton. Simulation results demonstrate that the BMA-CRM design is not only robust but also performing well.

Scenario		Benchmark						MTD selection percentage					
(MTD)	Model	1	2	3	4	5	6	1	2	3	4	5	6
1	1	0.509	0.192	0.239	0.049	0.011	0.001	0.588	0.020	0.391	0.001	0.001	0
(1)	2	0.345	0.452	0.153	0.017	0.033	0.002	0.350	0.573	0.018	0.059	0	0
	3	0.066	0.635	0.244	0.044	0.009	0.003	0.051	0.681	0.256	0.001	0.011	0
	4	0.027	0.225	0.637	0.061	0.049	0.002	0.001	0.202	0.758	0.040	0.000	0.001
	5	0.142	0.042	0.564	0.109	0.134	0.010	0.000	0.001	0.656	0.335	0.010	0
	6	0.497	0.120	0.050	0.186	0.110	0.039	0	0	0.008	0.759	0.234	0
	BMA	0.463	0.265	0.217	0.040	0.156	0.001	0.532	0.205	0.261	0.003	0	0
2	1	0.214	0.199	0.430	0.115	0.035	0.008	0.277	0.018	0.650	0.014	0.043	0
(2)	2	0.137	0.413	0.148	0.217	0.079	0.008	0.135	0.472	0.017	0.372	0.002	0.003
	3	0.026	0.368	0.401	0.113	0.069	0.024	0.011	0.379	0.475	0.016	0.120	0
	4	0.008	0.068	0.596	0.234	0.082	0.014	0	0.055	0.640	0.282	0.007	0.018
	5	0.035	0.011	0.196	0.590	0.135	0.036	0	0	0.189	0.712	0.099	0
	6	0.151	0.031	0.009	0.370	0.355	0.085	0	0	0	0.449	0.549	0.003
	BMA	0.167	0.208	0.429	0.138	0.052	0.009	0.199	0.166	0.510	0.102	0.023	0.002
3	1	0.117	0.138	0.428	0.148	0.135	0.036	0.145	0.007	0.616	0.035	0.195	0.004
(3)	2	0.073	0.247	0.161	0.383	0.110	0.027	0.070	0.296	0.014	0.570	0.013	0.039
	3	0.013	0.197	0.382	0.129	0.228	0.052	0.003	0.201	0.442	0.012	0.341	0.002
	4	0.008	0.032	0.370	0.414	0.111	0.067	0	0.023	0.390	0.478	0.013	0.096
	5	0.005	0.002	0.081	0.580	0.263	0.071	0	0	0.072	0.625	0.298	0.006
	6	0.042	0.009	0.002	0.199	0.649	0.1	0	0	0	0.196	0.771	0.034
	BMA	0.082	0.113	0.390	0.257	0.130	0.030	0.093	0.074	0.462	0.238	0.118	0.016
4	1	0.020	0.100	0.302	0.210	0.309	0.061	0.018	0.004	0.445	0.076	0.448	0.011
(4)	2	0.011	0.081	0.146	0.538	0.160	0.065	0.008	0.091	0.018	0.746	0.042	0.097
	3	0.007	0.052	0.262	0.189	0.429	0.062	0.001	0.049	0.331	0.027	0.586	0.008
	4	0.014	0.005	0.184	0.498	0.158	0.142	0	0.003	0.186	0.592	0.025	0.196
	5	0.002	0	0.025	0.489	0.435	0.081	0	0	0.019	0.482	0.491	0.009
	6	0.018	0	0.001	0.101	0.769	0.113	0	0	0	0.092	0.841	0.068
	BMA	0.014	0.040	0.232	0.386	0.272	0.058	0.010	0.013	0.255	0.401	0.282	0.040
5	1	0.009	0.083	0.092	0.088	0.384	0.346	0.006	0.001	0.118	0.030	0.591	0.255
(5)	2	0.003	0.036	0.118	0.237	0.167	0.440	0.002	0.032	0.004	0.339	0.054	0.571
	3	0.001	0.021	0.094	0.121	0.473	0.291	0	0.019	0.113	0.010	0.613	0.246
	4	0.003	0.002	0.065	0.223	0.165	0.543	0	0	0.065	0.270	0.017	0.648
	5	0.001	0	0.008	0.168	0.467	0.358	0	0	0.04	0.172	0.594	0.231
	6	0.001	0	0	0.019	0.430	0.551	0	0	0	0.014	0.441	0.545
	BMA	0.003	0.022	0.074	0.148	0.328	0.426	0.003	0.004	0.068	0.144	0.337	0.445
6	1	0.004	0.041	0.025	0.050	0.195	0.686	0.001	0	0.023	0.011	0.314	0.652
(6)	2	0.001	0.011	0.078	0.089	0.107	0.715	0	0.005	0.001	0.116	0.023	0.855
	3	0.001	0.005	0.038	0.102	0.263	0.593	0	0.004	0.031	0.006	0.339	0.621
	4	0.002	0.001	0.022	0.085	0.145	0.747	0	0	0.020	0.104	0.011	0.866
	5	0.001	0	0.003	0.062	0.258	0.679	0	0	0.001	0.062	0.334	0.603
	6	0.001	0	0	0.008	0.174	0.819	0	0	0	0.004	0.173	0.823
	BMA	0.001	0.008	0.025	0.058	0.170	0.739	0	0	0.015	0.038	0.147	0.802

Table 3.2: Simulated values of Benchmark and MTD selection percentage


Scenario 1
Density
N = 2000 Bandwidth = 0.07819
Scenario 2
Density
N = 2000 Bandwidth = 0.01958
Scenario 3
Density
N = 2000 Bandwidth = 0.0661
Scenario 4
Density
N = 2000 Bandwidth = 0.02435
CRM 1
CRM 2
CRM 3
CRM 4
CRM 5
CRM 6
BMA−CRM

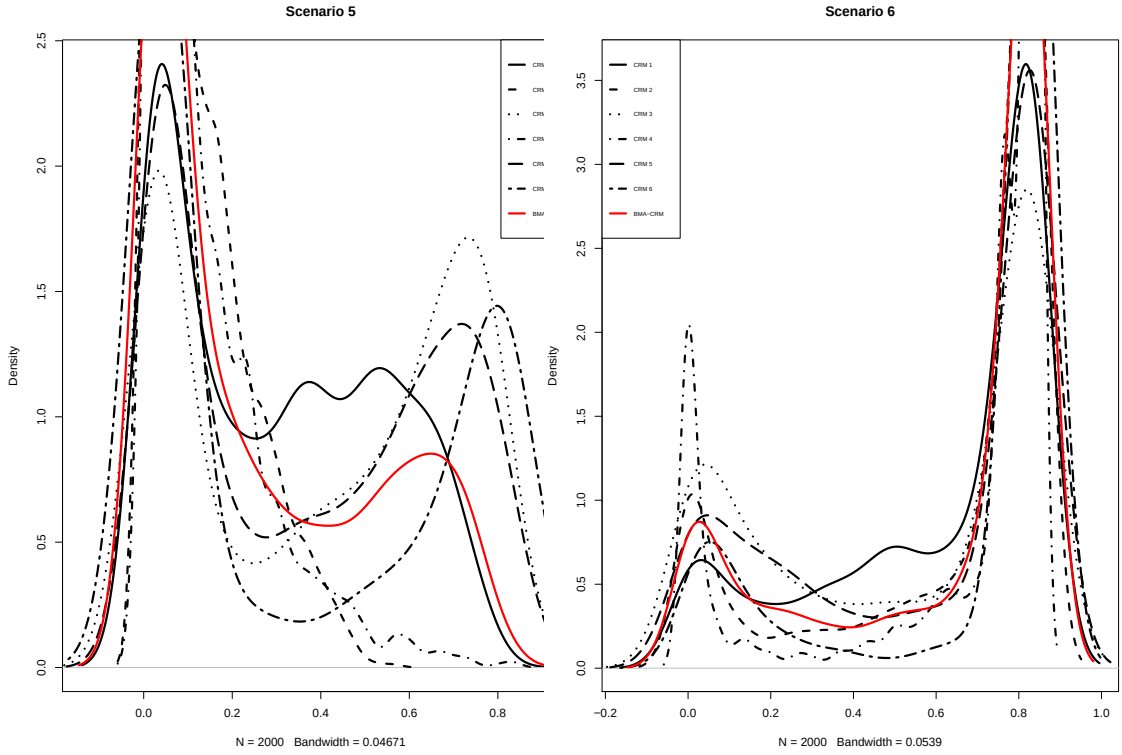


Figure 3.2: Comparison of density estimation of MTD allocation, scenario 1 to 6, no skipping

Scenario		Efficiency		Accuracy			Reliability		Safety	
(MTD)	Model	E1	E2 (s.d.)	A1	A2 (s.d.)	A3	R1	R2	S1	S2 (s.d.)
1	1	0	NAN (NAN)	0.588	0.482 (0.398)	0.531	0.395	0.405	0.413	0.104 (0.173)
(1)	2	0	NAN (NAN)	0.35	0.307 (0.389)	0.048	0.935	0.948	0.949	0.185 (0.226)
	3	0	NAN (NAN)	0.051	0.074 (0.175)	0.939	0.006	0	0.245	0.023 (0.065)
	4	0	NAN (NAN)	0	0.034 (0.005)	0	0.942	1	0.100	0.913 (0.213)
	5	0	NAN (NAN)	0	0.034 (0.006)	0	0.924	0.999	1	0.193(0.219)
	6	0	NAN (NAN)	0	0.033 (0)	0	0.937	1	1	0.913 (0.204)
	BMA	0	NAN (NAN)	0.532	0.405 (0.383)	0.431	0.45	0.47	0.469	0.119 (0.182)
2	1	0.277	0.249 (0.345)	0.018	0.126 (0.143)	0.028	0.578	0.724	0.706	0.156 (0.190)
(2)	2	0.135	0.154 (0.293)	0.472	0.375 (0.360)	0.41	0.362	0.499	0.393	0.118 (0.186)
	3	0.011	0.043 (0.087)	0.379	0.334 (0.392)	0.344	0.557	0.609	0.611	0.156 (0.214)
	4	0	0.033 (0.002)	0.055	0.079 (0.184)	0.053	0.869	0.943	0.946	0.222 (0.253)
	5	0	0.033(0.006)	0	0.034 (0.005)	0	0.921	1	1	0.233 (0.249)
	6	0	0.033 (0)	0	0.033 (0.002)	0	0.931	1	1	0.0233 (0.242)
	BMA	0.199	0.816 (0.309)	0.166	0.819 (0.258)	0.144	0.539	0.704	0.636	0.156 (0.203)
3	1	0.151	0.127 (0.218)	0.616	0.437 (0.336)	0.498	0.136	0.349	0.233	0.103 (0.156)
(3)	2	0.365	0.176(0.283)	0.014	0.117(0.134)	0.023	0.524	0.742	0.621	0.177 (0.224)
	3	0.204	0.122 (0.240)	0.442	0.339 (0.342)	0.371	0.331	0.526	0.355	0.139 (0.207)
	4	0.023	0.043 (0.089)	0.39	0.331 (0.380)	0.353	0.528	0.603	0.578	0.194 (0.241)
	5	0	0.033(0.001)	0.072	0.088 (0.197)	0.066	0.848	0.927	0.928	0.282 (0.276)
	6	0	0.033(0.003)	0	0.034 (0.004)	0	0.938	1	1	0.3 (0.259)
	BMA	0.167	0.120 (0.221)	0.462	0.346 (0.329)	0.358	0.277	0.469	0.371	0.138(0.197)
4	1	0.466	0.154 (0.199)	0.076	0.141 (0.124)	0.006	0.289	0.623	0.459	0.199 (0.209)
(4)	2	0.116	0.085 (0.150)	0.746	0.501(0.298)	0.591	0.114	0.223	0.138	0.121 (0.178)
	3	0.380	0.126(0.214)	0.027	0.133 (0.143)	0.034	0.529	0.696	0.594	0.245 (0.251)
	4	0.188	0.084 (0.179)	0.592	0.420 (0.324)	0.483	0.219	0.374	0.221	0.164 (0.233)
	5	0.019	0.039 (0.064)	0.482	0.372 (0.367)	0.407	0.482	0.513	0.500	0.256 (0.235)
	6	0	0.033 (0.003)	0.092	0.100 (0.208)	0.085	0.842	0.903	0.908	0.400 (0.257)
	BMA	0.277	0.115 (0.185)	0.401	0.298 (0.269)	0.272	0.25	0.444	0.322	0.178 (0.215)
5	1	0.155	0.083 (0.134)	0.591	0.301 (0.237)	0.247	0.337	0.380	0.255	0.367 (0.298)
(5)	2	0.376	0.113(0.170)	0.054	0.111 (0.106)	0.001	0.49	0.711	0.571	0.435 (0.352)
	3	0.142	0.075 (0.139)	0.613	0.408 (0.304)	0.467	0.263	0.345	0.246	0.292(0.312)
	4	0.335	0.095 (0.180)	0.017	0.116 (0.130)	0.019	0.601	0.734	0.648	0.505 (0.365)
	5	0.176	0.069 (0.148)	0.594	0.371 (0.295)	0.394	0.320	0.386	0.231	0.355 (0.311)
	6	0.014	0.036 (0.047)	0.441	0.329 (0.344)	0.370	0.579	0.548	0.545	0.525 (0.350)
	BMA	0.218	0.087 (0.147)	0.337	0.246 (0.249)	0.217	0.435	0.527	0.445	0.406 (0.344)
6	1	0.349	0.078 (0.112)	0.652	0.610 (0.274)	0.694	0	0.116	0	NAN (NAN)
(6)	2	0.145	0.070(0.111)	0.855	0.648 (0.284)	0.777	0	0.137	0	NAN (NAN)
	3	0.379	0.094 (0.147)	0.621	0.529 (0.332)	0.591	0	0.237	0	NAN (NAN)
	4	0.135	0.065(0.118)	0.866	0.677 (0.280)	0.824	0	0.135	0	NAN (NAN)
	5	0.397	0.083 (0.151)	0.603	0.584 (0.319)	0.659	0	0.184	0	NAN (NAN)
	6	0.177	0.069 (0.124)	0.823	0.698 (0.278)	0.821	0	0.131	0	NAN (NAN)
	BMA	0.199	0.073 (0.120)	0.802	0.636 (0.293)	0.747	0	0.143	0	NAN (NAN)

Table 3.3: Simulated values of EARS, allowing skipping.

Chapter 4

CRM without undue risk of toxicity

4.1 Introduction

Over the past years, statistical properties, important issues and CRM performance have been investigated, and many kinds of extensions in various directions have been proposed. The book by Cheung (2011) is a comprehensive evaluation of CRM. A review of CRM is provided by O'Quigley and Conaway (2010) who discuss both frequentist and Bayesian approaches, large- and small-sample properties, and some extensions. Garrett-Mayer (2006) provides a tutorial of CRM; Neuenschwander et al. (2008) elaborate on some critical aspects of the Bayesian approach to CRM; and Marchenko et al. (2013) provide an overview of several adaptive designs of phase I clinical trials, covering statistical, practical and logistical issues. Le Tourneau et al. (2009) discuss the pros and cons of different designs of phase I clinical trials including CRM. Recently, Clertant and O'Quigley (2017) and

Clertant and O'Quigley (2019) introduce semiparametric extensions of CRM. Iasonos et al. (2008) systematically compare the performance of CRM with nonparametric designs, and Iasonos and O'Quigley (2014) provide a comprehensive review of CRM and its extensions in medical practice. Sverdlov et al. (2014) comprehensively review various types of nonparametric and parametric designs, and discuss important issues such as overdose control, the Bayesian decision theoretical design and other optimal designs, and data analysis following adaptive designs.

All medical studies are inevitably challenged by the delicate issue of ethics. Particularly in any clinical trial of a new medical drug or intervention, we typically face conflict between individual ethics and collective ethics. Although the overarching goal of a phase I clinical trial is to identify the MTD and assign as many patients as possible to the MTD, ethical consideration dictates that we maximize the number of patients assigned to doses that have potential therapeutic or preventive benefits and at the same time prevent undue risk of toxicity to patients in the trial. This latter ethical requirement is the primary concern of this chapter, and in this chapter we propose a new CRM design which simultaneously minimizes undue risk of toxicity and maximizes efficiency. This is demonstrated by means of simulations in comparison with currently available CRM designs. In fact, through simulations we have observed that given any MTD and based on the skeleton suggested by the getprior() function from the dfcrm R package, no or few patients are assigned to any dose level above the MTD, and the majority of

patients are assigned to the MTD with our new CRM design.

This chapter proposes a CRM design with interesting properties and excellent performance. The performance of a phase I clinical trial design can be assessed by the same template consisting of several important and desirable performance criteria: BEARS (Benchmark, Efficiency, Accuracy, Reliability, Safety)

4.2 The new CRM design

We assume that there are K dose levels $d_1, d_2, \cdots, d_K$ to be investigated for their toxicity probabilities. Associated with the doses are some pre-specified skeleton probabilities $p_i, i = 1, 2, \cdots, K$, which are monotonically increasing in i. We wish to investigate the true toxicity probabilities $\pi_i = \pi_i(d_i) = \pi(p_i)$ at all dose levels linked by the skeleton. A common assumption is that π_i is an increasing function of p_i, and this assumption defines a partial order over the set $\{d_1, \cdots, d_k\}$ of all dose levels. In the CRM design, a form of the function π is assumed and involves unknown parameters. Different forms of the function π give rise to different CRM designs. To ensure safety of trial patients, a target DLT rate $\theta \in (0, 1)$ is specified and ethically acceptable (such as 0.2 in this paper). In any CRM design, the goal is to identify the MTD which is defined as the dose level d^* with true toxicity probability nearest to θ. That is, $d^* = arg\ min\{|\pi(p_i) - \theta|, i = 1, 2, \cdots, K\}$.

We introduce a new dose toxicity probability function $\pi_i = \frac{2\Phi(\beta+\alpha p_i)}{1+\Phi(\beta+\alpha p_i)}$, where Φ is the cumulative distribution function of the normal distribution $N(\mu, \sigma^2)$

and α is positive. The partial derivative of the probability π_i in p_i is given by $\frac{\partial \pi_i}{\partial p_i} = \frac{2\alpha\phi(\beta+\alpha p_i)}{(1+\Phi(\beta+\alpha p_i))^2} > 0$, where ϕ is the probability density function of $N(\mu, \sigma^2)$. This shows that the dose toxicity probability function is monotonically increasing in the dose level d_i for fixed values of other parameters.

The new CRM function involves the skeleton values $p_i, i = 1, 2, \cdots, K$, which are obtained from the getprior() function in the dfcrm R package, β, α, and (μ, σ^2) for the cumulative function Φ. When μ and other parameters are fixed, changing the value of σ^2 corresponds to stretching or compressing the function Φ, and it changes the dose toxicity probabilities when all other parameters are fixed. This represents a class of CRM designs and is an extra benefit of our new model. This is our advantage in the current paper. We fix all parameters β, α and $\mu = 0$ and allow the variance σ^2 to be random and follow a certain prior distribution. As shown in the next two sections, this new CRM design does not choose any dose that is above the MTD (except for when the MTD is the lowest dose), and hence avoids any potential overdosing. At the same time, the MTD is identified with a high probability. As a result, the risk of overdosing is eliminated, while the chance of better treatment at the MTD is maximized. This suggests that our new CRM design achieves both individual ethics and collective ethics simultaneously.

Our new CRM function $\pi(p_i) = \frac{2\Phi(\beta+\alpha p_i)}{1+\Phi(\beta+\alpha p_i)}$ monotonically increases in the dose level d_i (through p_i) when the unknown parameter σ^2 is fixed, and it is monotonic in the unknown parameter σ^2 at any fixed dose. Both monotone

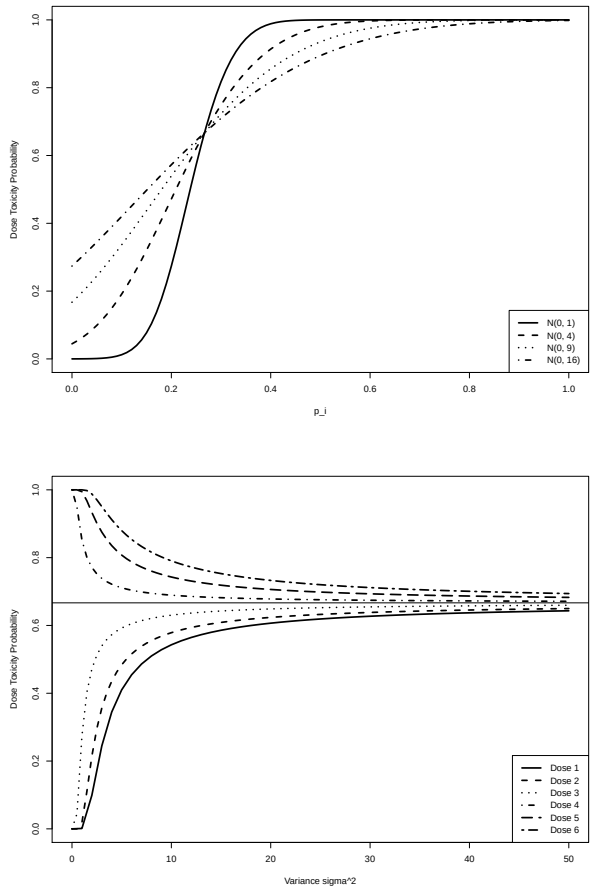


Figure 4.1: New CRM model $\pi(p_i) = \frac{2\Phi(-4+15p_i)}{1+\Phi(-4+15p_i)}$ as a function of p_i (top) and σ^2 (bottom)

properties are shown in Figure 4.1 for $\pi(p_i) = \frac{2\Phi(-4+15p_i)}{1+\Phi(-4+15p_i)}$. The figure also shows the effect of the value of σ^2 on the behavior of the new model. The six curves on the right correspond to taking p_i in the skeleton $(0.05, 0.11, 0.2, 0.31, 0.42, 0.53)$ which is scenario 5 in the next section. In all cases, we take $\mu = 0$.

	True Toxicity Probabilities						Skeleton						
Scenario	1	2	3	4	5	6	1	2	3	4	5	6	MTD
1	**0.200**	0.260	0.280	0.300	0.350	0.500	0.200	0.310	0.420	0.530	0.630	0.720	1
2	0.050	0.100	**0.200**	0.350	0.500	0.700	0.050	0.110	0.200	0.310	0.420	0.530	3
3	0.010	0.020	0.050	0.090	**0.180**	0.400	0.004	0.020	0.050	0.110	0.200	0.310	5
4	0.010	0.020	0.050	0.110	0.140	**0.210**	0.0004	0.004	0.020	0.050	0.110	0.200	6
5	0	0	**0.160**	0.300	0.350	0.400	0.050	0.110	0.200	0.310	0.420	0.530	3
6	0	0	0	**0.230**	0.300	0.350	0.020	0.050	0.110	0.200	0.310	0.420	4

Table 4.1: Summary of simulation scenarios

4.3 Simulation results and discussion

In this section, we introduce and discuss simulation results to compare the performance of our new model with the existing CRM models. There are six scenarios listed in Table 4.1 together with their respective MTDs. These scenarios are taken from Clertant and O'Quigley (2017), in which the pre-specified target DLT rate is $\theta = 0.2$. The dose level with a final estimated toxicity probability nearest to θ is set to be the estimated MTD. The parameter μ is 0, but the parameters α and β are values that offer good models. Each trial is simulated for 1,000 runs.

Each simulation run starts with treating the first cohort of patients at the lowest dose level. The cohort size is set to one. The decision of dose escalation or de-escalation is determined by the accumulated data at that time and the estimated toxicity probability curve. At the boundaries, dose escalation at d_K or de-escalation at d_1 does not occur. The size of escalation or de-escalation is determined by the dose level with an estimated toxicity probability nearest to the

Scenario		Skipping is not permitted						Skipping is allowed					
(MTD)	Model	1	2	3	4	5	6	1	2	3	4	5	6
1	I(a)	0.704	0.296	0	0	0	0	0.704	0.296	0	0	0	0
(1)	II	0.515	0.285	0.093	0.069	0.031	0.007	0.491	0.277	0.130	0.054	0.036	0.012
	III	0.315	0.247	0.186	0.138	0.092	0.022	0.363	0.254	0.186	0.141	0.040	0.016
	IV	**0.761**	0.148	0.060	0.025	0.006	0	**0.762**	0.147	0.059	0.026	0.006	0
2	I(b)	0.060	0.295	0.645	0	0	0	0.054	0.256	0.690	0	0	0
(3)	II	0.015	0	**0.758**	0.198	0.023	0.006	0.020	0	**0.739**	0.223	0.017	0.001
	III	0.018	0.114	0.543	0.238	0.085	0.002	0.031	0.130	0.560	0.234	0.042	0.003
	IV	0.111	0.273	0.404	0.190	0.022	0	0.122	0.267	0.416	0.173	0.022	0
3	I(b)	0.022	0.014	0.043	0.221	0.700	0	0.018	0.002	0.011	0.197	0.772	0
(5)	II	0.038	0	0	0	**0.819**	0.143	0.023	0	0	0	**0.834**	0.143
	III	0.083	0.001	0	0.093	0.572	0.251	0.046	0	0	0.096	0.650	0.208
	IV	0.063	0.005	0.040	0.220	0.539	0.133	0.06	0.002	0.038	0.225	0.572	0.103
4	I(c)	0.03	0.005	0.024	0.065	0.193	0.683	0.027	0	0.011	0.082	0.189	0.691
(6)	II	0.031	0.001	0	0	0	**0.968**	0.01	0	0	0	0	**0.990**
	III	0.018	0	0	0.003	0.053	0.926	0.011	0	0	0.001	0.060	0.928
	IV	0.039	0.007	0.022	0.121	0.211	0.600	0.039	0.003	0.013	0.099	0.276	0.570
5	I(b)	0.026	0	**0.974**	0	0	0	0.026	0	**0.974**	0	0	0
(3)	II	0.019	0	0.525	0.313	0.088	0.055	0.011	0	0.534	0.306	0.101	0.048
	III	0.020	0	0.380	0.284	0.221	0.095	0.026	0	0.432	0.301	0.137	0.104
	IV	0.053	0	0.532	0.294	0.102	0.019	0.049	0	0.527	0.303	0.104	0.017
6	I(b)	0.091	0	0	**0.909**	0	0	0.093	0	0	**0.907**	0	0
(4)	II	0.023	0	0	0.581	0.268	0.128	0.006	0	0	0.528	0.272	0.194
	III	0.055	0	0	0.398	0.299	0.248	0.051	0	0	0.440	0.268	0.241
	IV	0.093	0	0	0.556	0.260	0.091	0.104	0	0	0.549	0.246	0.101

Table 4.2: Simulated values of Benchmark, with or without skipping. Bold values are the highest in each group.

target DLT rate and by whether skipping is permitted. Escalation or de-escalation occurs one step away from the current dose at most if skipping is not allowed.

Suppose that $m-1$ patients have been treated and

$$D_m = \{(d^{(1)}, y_1), \cdots, (d^{(m-1)}, y_{m-1})\}$$

is the observed information, where $d^{(i)}$ is the dose level applied to patient $i = 1, \cdots, m-1$, and y_i is the patient toxicity response (1 for toxicity and 0 otherwise). Suppose that $\hat{\pi}_m(p_i)$ is the estimated toxicity probability at dose level d_i when the m^{th} patient is to be assigned a dose level. Suppose that the index j minimizes $|\hat{\pi}_m(p_i) - \theta|$ over $i \in \{1, 2, \cdots, K\}$. If $d_j = d^{(m-1)}$, then the dose level $d^{(m-1)}$ is allocated. If $d_j > d^{(m-1)}$, the dose level d_j is assigned if skipping is allowed, but the dose level $d^{(m-1)} + 1$ is assigned if skipping is not allowed and is subject to the boundary rule. Similarly, if $d_j < d^{(m-1)}$, the dose level d_j is assigned if skipping is allowed, but the dose level $d^{(m-1)} - 1$ is assigned if skipping is not allowed and is subject to the boundary rule. This process continues until a pre-specified number of patients are treated, which is 30 in this paper. At the conclusion of the trial, the final toxicity probabilities $\hat{\pi}(p_i), i = 1, 2 \cdots, K$, are updated, and the dose level with an estimated toxicity probability nearest to the target θ is chosen as the MTD.

We follow the Bayesian approach to estimate $\hat{\pi}_m(p_i)$. After calculating the likelihood function, the prior distribution is updated into a posterior distribution, and the estimator $\hat{\pi}_m(p)$ is given by the posterior mean toxicity probability of

$\pi(p)$ after treating $m-1$ patients. The likelihood function is given by

$$L(\sigma^2|D_m) = \prod_{i=1}^{m-1} \left\{ \frac{2\Phi(\beta + \alpha p^{(i)})}{1 + \Phi(\beta + \alpha p^{(i)})} \right\}^{y_i} \left\{ 1 - \frac{2\Phi(\beta + \alpha p^{(i)})}{1 + \Phi(\beta + \alpha p^{(i)})} \right\}^{1-y_i},$$

where $p^{(i)}$ is the skeleton value associated with dose $d^{(i)}$. We assume that the positive parameter σ^2 follows a prior distribution $f(\sigma^2)$. The posterior mean toxicity probability $\hat{\pi}_m(p_i)$ at dose level d_i is estimated to be

$$\hat{\pi}_m(p_i) = \int \frac{2\Phi(\beta + \alpha p_i)}{1 + \Phi(\beta + \alpha p_i)} \frac{L(\sigma^2|D_m) f(\sigma^2)}{\int L(\sigma^2|D_m) f(\sigma^2) d(\sigma^2)} d(\sigma^2).$$

We simulate each CRM model under the same conditions; these conditions include the same prior distribution of the unknown parameters and the same seed for each simulation run. The prior distribution is the gamma distribution $gamma(x, 0.5, 0.5)$ from R software.

We systematically compare the performance of our new design against the existing ones in the important BEARS operating criteria: Benchmark, Efficacy, Accuracy, Reliability, Safety. The Benchmark criterion is introduced by Cheung (2014) and ARS criteria are introduced by Zhou et al. (2018). We now introduce the Efficacy criterion. Together, they serve as a template of BEARS for choosing a CRM design that satisfies both individual ethics and collective ethics.

To save space in the tables, our models are summarized as follows, where p represents the skeleton value at each dose, and x represents the unknown parameter. The new models are $I(a) = \frac{2\Phi(-2+5p)}{1+\Phi(-2+5p)}$, $I(b) = \frac{2\Phi(-4+15p)}{1+\Phi(-4+15p)}$ and $I(c) = \frac{2\Phi(-4+3p)}{1+\Phi(-4+3p)}$, where the cumulative distribution function Φ is given by the normal distribution

Scenario		Efficiency		Accuracy			Reliability		Safety	
(MTD)	Model	E1	E2 (s.d.)	A1	A2 (s.d.)	A3	R1	R2	S1	S2 (s.d.)
1	I(a)	0	NAN (NAN)	0.741	0.702 (0.363)	0.716	0.278	0.162	0.259	0.060 (0.162)
(1)	II	0	NAN (NAN)	0.600	0.451 (0.355)	0.450	0.357	0.338	0.400	0.110 (0.161)
	III	0	NAN (NAN)	0.387	0.288 (0.301)	0.263	0.131	0.524	0.613	0.142 (0.129)
	IV	0	NAN (NAN)	**0.752**	**0.810** (0.268)	**0.824**	**0.031**	**0**	**0.248**	**0.038** (0.091)
2	I(b)	0.341	0.204 (0.247)	0.659	0.592 (0.384)	0.646	**0**	0.263	**0**	**0** (0)
(3)	II	**0**	**0.033** (0)	**0.855**	**0.621** (0.263)	**0.698**	0.167	**0.077**	0.145	0.104 (0.136)
	III	0.167	0.081 (0.128)	0.65	0.451 (0.193)	0.414	0.049	0.087	0.183	0.129 (0.102)
	IV	0.312	0.250 (0.154)	0.495	0.307 (0.144)	0.065	0.023	0.161	0.193	0.064 (0.095)
3	I(b)	0.282	0.101 (0.160)	0.718	0.597 (0.352)	**0.697**	**0**	0.234	**0**	**0** (0)
(5)	II	**0**	**0.033** (0)	**0.872**	**0.600** (0.223)	0.681	0.150	**0.051**	0.128	0.267 (0.223)
	III	0.158	0.052 (0.076)	0.560	0.344 (0.194)	0.224	0.389	0.213	0.282	0.449 (0.238)
	IV	0.260	0.116 (0.076)	0.554	0.336 (0.134)	0.092	0.075	0.092	0.186	0.199 (0.177)
4	I(c)	0.317	0.075 (0.066)	0.683	0.626 (0.229)	0.706	0	0.044	0	NAN (NAN)
(6)	II	**0**	**0.033** (0)	**1**	**0.833** (0)	**1**	0	**0**	0	NAN (NAN)
	III	0.072	0.039 (0.046)	0.928	0.804 (0.106)	0.956	0	0.001	0	NAN (NAN)
	IV	0.331	0.105 (0.077)	0.669	0.474 (0.244)	0.526	0	0.145	0	NAN (NAN)
5	I(b)	0.009	0.072 (0.075)	**0.991**	**0.856** (0.106)	**0.982**	**0**	**0**	**0**	**0** (0)
(3)	II	**0**	**0.033** (0)	0.636	0.430 (0.296)	0.425	0.347	0.264	0.364	0.168 (0.168)
	III	0.034	0.047 (0.063)	0.478	0.305 (0.235)	0.240	0.176	0.365	0.488	0.200 (0.171)
	IV	0.105	0.151 (0.084)	0.448	0.339 (0.156)	0.119	0.052	0.152	0.447	0.119 (0.120)
6	I(b)	0.057	0.080 (0.082)	**0.943**	**0.760** (0.141)	**0.915**	**0**	**0**	**0**	**0** (0)
(4)	II	**0**	**0.033** (0)	0.735	0.472 (0.3)	0.501	0.312	0.229	0.265	0.214 (0.210)
	III	0.086	0.051 (0.071)	0.476	0.278 (0.219)	0.183	0.325	0.381	0.438	0.284 (0.222)
	IV	0.157	0.126 (0.082)	0.534	0.328 (0.146)	0.085	0.070	0.150	0.309	0.147 (0.157)

Table 4.3: Simulated values of EARS, not allowing skipping. Bold values are the best in each group.

Scenario		Efficiency		Accuracy			Reliability		Safety	
(MTD)	Model	E1	E2 (s.d.)	A1	A2 (s.d.)	A3	R1	R2	S1	S2 (s.d.)
1	I(a)	0	NAN (NAN)	0.741	0.702 (0.363)	0.716	0.278	0.162	0.259	0.060 (0.162)
(1)	II	0	NAN (NAN)	0.589	0.437 (0.352)	0.439	0.367	0.346	0.411	0.113 (0.160)
	III	0	NAN (NAN)	0.421	0.315 (0.312)	0.309	0.140	0.483	0.579	0.137 (0.129)
	IV	0	NAN (NAN)	**0.751**	**0.811** (0.266)	**0.823**	**0.029**	**0**	**0.249**	**0.038** (0.091)
2	I(b)	0.305	0.182 (0.244)	0.695	**0.637** (0.374)	0.680	**0**	0.237	**0**	**0** (0)
(3)	II	**0**	**0.017** (0)	**0.853**	0.612 (0.259)	**0.689**	0.194	**0.073**	0.147	0.118 (0.139)
	III	0.208	0.082 (0.152)	0.633	0.478 (0.207)	0.470	0.074	0.083	0.159	0.119 (0.106)
	IV	0.308	0.254 (0.150)	0.497	0.302 (0.142)	0.055	0.027	0.164	0.195	0.064 (0.095)
3	I(b)	0.222	0.074 (0.155)	0.778	**0.705** (0.343)	0.756	**0**	0.170	**0**	**0** (0)
(5)	II	**0**	**0.008** (0)	**0.886**	0.658 (0.224)	**0.768**	0.179	**0.040**	0.114	0.308 (0.224)
	III	0.149	0.032 (0.089)	0.605	0.401 (0.214)	0.340	0.395	0.151	0.246	0.469 (0.26)
	IV	0.275	0.112 (0.076)	0.544	0.367 (0.137)	0.134	0.066	0.071	0.181	0.186 (0.175)
4	I(c)	0.317	0.065 (0.072)	0.683	0.675 (0.257)	0.715	0	0.022	0	NAN (NAN)
(6)	II	**0**	**0.007** (0)	**1**	**0.967** (0)	**1**	0	**0**	0	NAN (NAN)
	III	0.073	0.014 (0.055)	0.927	0.928 (0.125)	0.971	0	0.001	0	NAN (NAN)
	IV	0.363	0.100 (0.079)	0.637	0.498 (0.261)	0.505	0	0.126	0	NAN (NAN)
5	I(b)	0.010	0.065 (0.083)	**0.990**	**0.870** (0.117)	**0.982**	**0**	**0**	**0**	**0** (0)
(3)	II	**0**	**0.017** (0)	0.668	0.426 (0.304)	0.455	0.348	0.249	0.332	0.177 (0.173)
	III	0.025	0.032 (0.067)	0.516	0.329 (0.265)	0.304	0.232	0.338	0.459	0.203 (0.182)
	IV	0.106	0.155 (0.078)	0.443	0.326 (0.151)	0.093	0.061	0.168	0.451	0.122 (0.121)
6	I(b)	0.049	0.071 (0.090)	**0.951**	**0.788** (0.156)	**0.917**	**0**	**0**	**0**	**0** (0)
(4)	II	**0**	**0.011** (0)	0.758	0.487 (0.312)	0.538	0.312	0.227	0.242	0.240 (0.218)
	III	0.101	0.032 (0.079)	0.486	0.313 (0.259)	0.271	0.366	0.364	0.413	0.296 (0.246)
	IV	0.162	0.125 (0.078)	0.526	0.330 (0.140)	0.070	0.081	0.140	0.312	0.147 (0.157)

Table 4.4: Simulated values of EARS, allowing skipping. Bold values are the best in each group.

Scenario		Skipping is not permitted						Skipping is allowed					
(MTD)	Model	1	2	3	4	5	6	1	2	3	4	5	6
1	I(a)	0.741	0.259	0	0	0	0	0.741	0.259	0	0	0	0
(1)	II	0.600	0.252	0.081	0.048	0.019	0	0.589	0.269	0.076	0.046	0.019	0.001
	III	0.387	0.338	0.143	0.072	0.032	0.028	0.421	0.313	0.153	0.069	0.026	0.018
	IV	**0.752**	0.157	0.056	0.030	0.005	0	**0.751**	0.158	0.056	0.030	0.005	0
2	I(b)	0.011	0.330	0.650	0	0	0	0.013	0.292	0.695	0	0	0
(3)	II	0	0	**0.855**	0.142	0.003	0	0	0	**0.853**	0.146	0.001	0
	III	0	0.167	0.650	0.158	0.023	0.002	0.001	0.207	0.633	0.139	0.017	0.003
	IV	0.068	0.244	0.495	0.179	0.014	0	0.068	0.240	0.497	0.178	0.017	0
3	I(b)	0	0	0.008	0.274	0.718	0	0	0.001	0.005	0.216	0.778	0
(5)	II	0	0	0	0	**0.872**	0.128	0	0	0	0	**0.886**	0.114
	III	0	0	0	0.158	0.560	0.282	0	0	0	0.149	0.605	0.246
	IV	0	0.003	0.038	0.219	0.554	0.186	0	0.002	0.045	0.228	0.544	0.181
4	I(c)	0.009	0.013	0.028	0.052	0.215	0.683	0.030	0.012	0.030	0.043	0.202	0.683
(6)	II	0	0	0	0	0	**1**	0	0	0	0	0	**1**
	III	0	0	0	0.004	0.068	0.928	0	0	0	0.001	0.072	0.927
	IV	0.003	0.007	0.026	0.079	0.216	0.669	0.007	0.002	0.037	0.084	0.233	0.637
5	I(b)	0	0.009	**0.991**	0	0	0	0	0.01	**0.99**	0	0	0
(3)	II	0	0	0.636	0.317	0.038	0.009	0	0	0.668	0.289	0.036	0.007
	III	0	0.034	0.478	0.319	0.094	0.075	0	0.025	0.516	0.308	0.077	0.074
	IV	0	0.105	0.448	0.351	0.08	0.016	0	0.106	0.443	0.35	0.084	0.017
6	I(b)	0	0	0.057	**0.943**	0	0	0	0	0.049	**0.951**	0	0
(4)	II	0	0	0	0.735	0.222	0.043	0	0	0	0.758	0.210	0.032
	III	0	0	0.086	0.476	0.221	0.217	0	0	0.101	0.486	0.224	0.189
	IV	0	0	0.157	0.534	0.224	0.085	0	0	0.162	0.526	0.238	0.074

Table 4.5: Simulated percentages of MTD selection, with or without skipping. Bold values are the highest in each group.

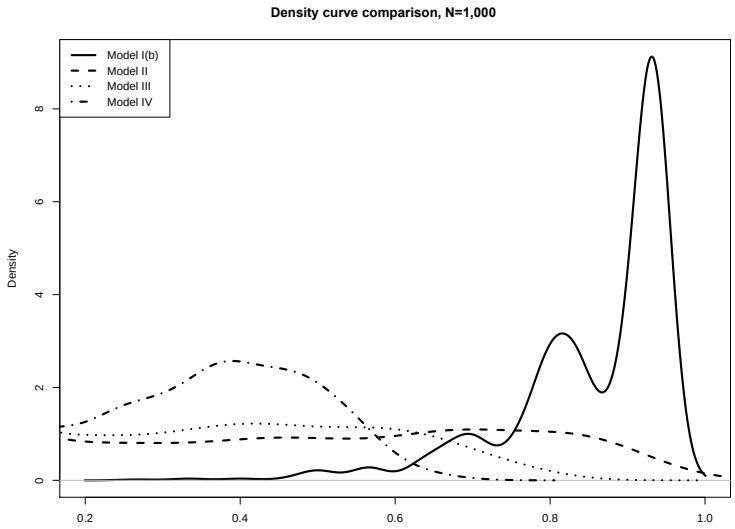


Figure 4.2: Comparison of density estimation of MTD allocation, scenario 5, no skipping.

$N(0, x)$. The power model is $II = p^{exp(x)}$, the logistic model is $III = \frac{exp(-2+xp)}{1+exp(-2+xp)}$, and the hyperbolic tangent model is $IV = \left(\frac{exp(p)}{exp(p)+exp(-p)}\right)^x$.

We summarize some key observations as follows to demonstrate that our new CRM design achieves individual and collective ethics simultaneously. Table 4.2 gives simulated values of the benchmark, and Table 4.3 and Table 4.4 (which are similar) give simulated values of EARS.

Benchmark means using sample mean toxicity probabilities to choose the MTD.

- **Performance measure:** The Benchmark provides the estimated distribution of the MTD, which is determined by the dose level with an observed

sample mean toxicity probability nearest to the target θ. The higher the proportion of correct selection of the MTD, the better the design.

- **Results:** We see from Table 4.2 that our new model performs excellent overall, especially when all doses below the MTD have 0 toxicity probabilities. In other cases, the Benchmark of the new model is not far behind that of the best model.

Efficiency checks the avoidance of assigning patients to ineffective doses.We would like to treat as many patients effectively as possible. This means treating the least patients as possible on any dose lower than the MTD.

- **Two performance measures:** The Efficacy measure E1 calculates the proportion of the simulation runs that assign patients to a dose lower than the MTD. The lower the proportion E1, the better the design. E2 reports the mean and standard deviation of the percentage of patients assigned to a dose lower than the MTD. The lower the mean, the better the design.

- **Results:** From Table 4.3 and Tables 4.4, overall, our new design performs well on Efficacy. Although our design assigns patients to either the MTD or lower doses, it is highly efficient because the values of E1 and E2 are reasonably low.

Accuracy depicts the assignment of patients to the MTD.

- **Three performance measures:** A1 measures the proportion of the simulation runs that identify the MTD. The higher the proportion, the better the design. A2 calculates the mean and standard deviation of the percentage of patients assigned to the MTD. The higher the mean, the better the design. A3 gives the proportion of the simulation runs that assign more than 50% of patients to the MTD. The higher the proportion, the better the design.

- **Results:** From Table 4.3 and Tables 4.4, overall, our new model is accurate. All three measures are reasonably large in most cases.

Reliability concerns the risk of severe overdosing.

- **Two performance measures:** R1 measures the proportion of the simulation runs that allocate more than 50% of patients to a dose higher than the MTD. The lower the proportion, the better the design. R2 calculates the proportion of the simulation runs with less than one-sixth (i.e., 5 in our simulations) of patients at the MTD. The lower the proportion, the better the design.

- **Results:** From Table 4.3 and Tables 4.4, our new model is the most reliable overall. Not considering the boundary cases, our new design assigns no patients to a dose greater than the MTD. Further, the value of R2 is reasonably low.

Safety refers to the protection from overdosing.

- **Two performance measures:** S1 measures the proportion of the simulation runs that result in doses greater than the MTD. The lower its value, the better the design. S2 calculates the mean and standard deviation of the percentage of patients assigned to a dose greater than the MTD. The lower the mean, the better the design.

- **Results:** From Table 4.3 and Tables 4.4, our model is the safest in almost all cases! With our new design and ignoring the boundary cases, there are no patients assigned to a dose greater than the MTD. In contrast, the power model assigns patients to either the MTD or doses greater than the MTD.

The most distinguishing feature is that our new CRM design assigns no patients to doses above the MTD, unless the MTD is the lowest dose. This is not the case for any other design. In fact, the power model performs the opposite way by assigning all patients to either the MTD or doses greater than the MTD. Avoiding doses greater than the MTD is important for individual ethics to mitigate potential harm of overdosing patients in the trial. Particularly when doses below the MTD are completely non-toxic (i.e., with 0 toxicity probabilities), our new design not only finds the MTD with almost certainty but also allocates almost all patients to the MTD. The performance is truly remarkable. This means that our new CRM design avoids the harm of overdosing and achieves an excellent balance between individual ethics and collective ethics.

To further understand the advantage of our new CRM design, Figure 4.2

shows the comparison of estimated density curves of the proportion of patients allocated to the MTD based on 1,000 simulation runs. Our new design offers the highest proportion of patients assigned to the MTD.

4.4 Conclusions

Although our new CRM design has the advantage of assigning zero patients to overly toxic doses, it has the disadvantage of having only isolated information at or below the MTD. This may lead to a weakness of estimating the overall dose toxicity probability curve. However, as pointed out by O'Quigley and Conaway (2010), such estimation is rarely a goal of the phase I clinical trial.

There is a famous saying that "All models are wrong, but some are useful" (George Box). Our proposed model is not perfect, but it is useful to guide dose escalation and de-escalation decisions in practical situations of treating patients in a phase I clinical trial. We have identified functions with which the CRM design completely avoids potential overdosing. However, for each scenario, if we change the values of the parameters α and β, we could also observe different results. Nevertheless, our model serves a purpose if we can identify good models.

Chapter 5

The CRM design with combination drugs

5.1 Introduction

Recent years have seen significant interest and progress on personalized medicine, molecularly targeted therapies and combination drugs. For cytotoxic treatment of cancer, one particular drug may show effectiveness to destroy the cancerous cell, but cellular heterogeneity may create a certain drug resistant disease (Marusyk et al., 2012; Harrington et al., 2013). Because drug susceptibility varies among cells and between patients, a combination of drugs can help achieve the desirable treatment intensity and resistance, when the drugs have non-overlapping toxicities and hence the combination is not overly toxic (Harrington et al., 2013; Dancey and Chen, 2006). For example, combination drugs are shown to be effective in enhancing survival in early stage and advanced stage cancer patients, and even curative in testicular cancer patients (Harrington et al., 2013; Master and Köberle,

2003).

When a combination of drugs is used, it is important to determine the maximum tolerated dose combination for all drugs involved. This is not a simple matter because the dose toxicity probability model may be significantly influenced by the potential pharmacokinetic and pharmacodynamic interactions between the drugs (Harrington et al., 2013). Furthermore, due to a two dimensional lattice structure of dose combination levels, dose escalation and deescalation rules depend on the partial order of the combination drug levels. Kramar et al. (1999) propose the use of the continual reassessment method when the toxicity probabilities of the dose combination can be ordered a priori. Wages et al. (2011) use the continual reassessment method to estimate the MTD combination when the ordering of the toxicity probabilities cannot be known a priori. Wages and Conaway (2013) discuss the issues with choosing partial ordering for the combination drugs when the power model is used with the continual reassessment method. Riviere et al. (2015) compare the performance of different designs of Phase I clinical trials for combination drugs, including the up-and-down designs, the continual reassessment method with partial ordering, copula regression and the latent contingency table. Further improvements of these methods are suggested in Yin et al. (2015) and Wages (2015). Diniz et al. (2017) consider the continual reassessment method under various model scenarios and misspecification. On the other hand, Yin and Yuan (2009a) and Yuan and Yin (2011) use copula to describe stochastic dependence in combination drug trials, and Bailey et al. (2009) apply logistic

regression with covariates for the Bayesian design of phase I trial for combination drugs.

Hamberg et al. (2010) discuss important issues and pitfalls in designing phase I trials for combination drugs as well as methods to avoid bias due to imbalance in observed background toxicity. Hirakawa et al. (2014) compare various model-based designs, including the continual reassessment method and copula-based design, for phase I combination drug trials. The Clinical Trial Design Task Force offers useful recommendations on designing phase I trials for combination drugs (Paller et al., 2014). Wages et al. (2016) provide an excellent review on three methods for designing phase I trials of drug combinations and investigate their operating characteristics. In an excellent survey, Sverdlov et al. (2014) provide significant insight into the design of phase I drug combination trials. Sweeting and Pander (2012) investigate the performance of different escalation strategies for Phase I combination drug trials, and show that strategies allowing only non-diagonal escalations are inefficient and identify fewer MTD combinations.

To improve, in my thesis, I set up a complete order of drug movement ahead of time. we allow both diagonal and non-diagonal movements of dose escalation and de-escalation based on the complete order. Furthermore, we introduce two new complete orders.

5.2 Model-Based Designs for Combination Drugs

We extend the CRM model to combination drugs. Patients are treated sequentially, one at a time. We consider a combination of two drugs A and B. Drug A has K dose levels and drug B has L dose levels. At dose combination (i, j), $i = 1, 2 \cdots, K$ and $j = 1, 2, \cdots, L$. The pre-specified toxicity probability is given by p_{ij}. We introduce a new CRM design with the following new model of dose toxicity probability:

$$P(\text{toxicity at combination dose levels } (i,j)) = \pi(\alpha) = \frac{2\Phi(\eta + \alpha p_{ij})}{1 + \Phi(\eta + \alpha p_{ij})},$$

where η is taken as a constant, and α is an unknown parameter, and Φ is the cumulative distribution function of the normal distribution $N(\mu, \sigma^2)$ with given values of μ and σ^2.

For the purpose of simulation, we assume that the parameter α is random but follows some prior distribution. Different values of μ and σ^2 can be set in the simulation study, however from our simulation experience and our previous research, the standard normal distribution $N(0, 1)$ with $\mu = 0$ and $\sigma^2 = 1$ is a very good choice. The Bayesian approach is applied to estimate the prior or posterior mean toxicity probability at each combination of dose levels $(i, j), i = 1, \cdots, K; j = 1, \cdots, L$. Suppose patient $k, k = 1, 2, \cdots, n$, is treated at dose combination $(i^{(k)}, j^{(k)})$ and toxicity outcome y_k is observed, where $y_k = 1$ if the patient is toxic and $y_k = 0$ otherwise. Let $D_n = \{(i^{(1)}, j^{(1)}), y_1; (i^{(2)}, j^{(2)}), y_2; \cdots; (i^{(n)}, j^{(n)}), y_n\}$ be the observed information on dose combination and its corresponding toxicity of all previously

treated patients. Denote

$$\pi_k = \pi(\alpha) = \frac{2\Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})}{1 + \Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})},$$

Where π_k is the toxicity probability at the treated patient k, $k = 1, 2, \cdots, n$.

The likelihood function becomes

$$L(\alpha|D_n) = \prod_{k=1}^{n} \{\pi_k\}^{y_k} \{1 - \pi_k\}^{1-y_k}.$$

With our new model, this likelihood function is

$$L(\alpha|D_n) = \prod_{k=1}^{n} \left\{ \frac{2\Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})}{1 + \Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})} \right\}^{y_k} \left\{ 1 - \frac{2\Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})}{1 + \Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})} \right\}^{1-y_k}.$$

Suppose that α is positive and follows the prior distribution $f(\alpha)$. By the Bayes' Theorem, after treating n patients, the posterior mean toxicity probability at dose combination (i, j) is estimated to be

$$\hat{\pi}(i, j) = \int \frac{2\Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})}{1 + \Phi(\eta + \alpha p_{i^{(k)},\, j^{(k)}})} \frac{L(\alpha|D_n) f(\alpha)}{\int L(\alpha|D_n) f(\alpha) d\alpha} d\alpha.$$

The dose escalation or de-escalation procedure is sequential and the patient cohort is one. After treating each patient, we collect the toxicity data and calculate the posterior mean of toxicity probability at each combination of dose levels, say $\hat{\pi}(i, j), i = 1, \cdots, K; j = 1, \cdots, L$. The dose combination whose toxicity probability is closest to the target toxicity probability θ is recommended to the next patient. This target toxicity probability θ is regarded as the proportion of trial patients permitted to experience the DLT, and is often set to be 0.33.

The trial terminates when the sequence of toxicity probabilities converges or the maximum sample size of patients is reached. MTD of the combination drug is identified as the dose combination whose last updated posterior mean toxicity probability is nearest to the target toxicity probability θ.

5.3 Possible dose-toxicity orders

The entire procedure of dose-finding trials with a single drug depends on the monotonic relationship between dose and toxicity probability. In this case, each dose-toxicity probabilities can be easily ordered. But for a combination of two drugs, the ordering of toxicity probabilities of dose combination is unknown. For example, for 3×3 combination drug dose levels, the toxicity probability of dose combination $(3, 2)$ is greater than that of dose combination $(2, 2)$, meanwhile, we know the toxicity probability of dose combination $(2, 3)$ is greater than that of dose combination $(2, 2)$. But we may not know whether the toxicity probability of dose combination $(3, 2)$ is greater than that of dose combination $(2, 3)$. So specifying a possible dose-toxicity ordering is a primary goal of CRM with combination drugs.

Suppose for a combination-drug trial, Drug A has 4 dose levels and Drug B has 3 dose levels. There is a total of 12 combinations, such as $d_{11},\ d_{12},\ d_{13},\ \cdots,\ d_{42},\ d_{43}$, where d_{ij} represents dose combination (i, j), $i = 1, 2, 3, 4$ and $j = 1, 2, 3$. The matrix of combination doses is described in Table 5.1

In order to implement CRM, it is important to define a complete order among

Doses of	B			Simulation		
A	1	2	3	1	2	3
4	d_{41}	d_{42}	d_{43}	0.600	0.700	0.800
3	d_{31}	d_{32}	d_{33}	0.380	0.450	0.520
2	d_{21}	d_{22}	d_{23}	0.100	**0.200**	0.310
1	d_{11}	d_{12}	d_{13}	0.010	0.040	0.070

Table 5.1: Combination dose levels of Drug A and B. Combination dose of $(2, 2)$ (say d_{22}) is the MTD in the simulation.

$d_{11},\ d_{12},\ d_{13},\ \cdots,\ d_{42},\ d_{43}$. This can be done in different ways, depending on the movement of escalation and de-escalation. For example, we can move across rows, up columns, down columns or along diagonals. The following 6 possible orderings of dose combinations are introduced by Wages and Conaway (2013).

- $order = 1$: Across rows

$$d_{11} \leq d_{12} \leq d_{13} \leq d_{21} \leq d_{22} \leq d_{23} \leq d_{31} \leq d_{32} \leq d_{33} \leq d_{41} \leq d_{42} \leq d_{43}$$

- $order = 2$: Up columns

$$d_{11} \leq d_{21} \leq d_{31} \leq d_{41} \leq d_{12} \leq d_{22} \leq d_{32} \leq d_{42} \leq d_{13} \leq d_{23} \leq d_{33} \leq d_{43}$$

- $order = 3$: Up diagonals

$$d_{11} \leq d_{12} \leq d_{21} \leq d_{13} \leq d_{22} \leq d_{31} \leq d_{23} \leq d_{32} \leq d_{41} \leq d_{33} \leq d_{42} \leq d_{43}$$

- $order = 4$: Down diagonals

$$d_{11} \leq d_{21} \leq d_{12} \leq d_{31} \leq d_{22} \leq d_{13} \leq d_{41} \leq d_{32} \leq d_{23} \leq d_{42} \leq d_{33} \leq d_{43}$$

order \ location	1	2	3	4	5	6	7	8	9	10	11	12
1	d_{11}	d_{12}	d_{13}	d_{21}	$\mathbf{d_{22}}$	d_{23}	d_{31}	d_{32}	d_{33}	d_{41}	d_{42}	d_{43}
2	d_{11}	d_{21}	d_{31}	d_{41}	d_{12}	$\mathbf{d_{22}}$	d_{32}	d_{42}	d_{13}	d_{23}	d_{33}	d_{43}
3	d_{11}	d_{12}	d_{21}	d_{13}	$\mathbf{d_{22}}$	d_{31}	d_{23}	d_{32}	d_{41}	d_{33}	d_{42}	d_{43}
4	d_{11}	d_{21}	d_{12}	d_{31}	$\mathbf{d_{22}}$	d_{13}	d_{41}	d_{32}	d_{23}	d_{42}	d_{33}	d_{43}
5	d_{11}	d_{12}	d_{21}	d_{31}	$\mathbf{d_{22}}$	d_{13}	d_{23}	d_{32}	d_{41}	d_{42}	d_{33}	d_{43}
6	d_{11}	d_{21}	d_{12}	d_{13}	$\mathbf{d_{22}}$	d_{31}	d_{41}	d_{32}	d_{23}	d_{33}	d_{42}	d_{43}
7	d_{11}	d_{12}	d_{21}	d_{31}	d_{13}	$\mathbf{d_{22}}$	d_{23}	d_{41}	d_{32}	d_{42}	d_{33}	d_{43}
8	d_{11}	d_{21}	d_{12}	d_{13}	d_{31}	$\mathbf{d_{22}}$	d_{41}	d_{23}	d_{32}	d_{33}	d_{42}	d_{43}

Table 5.2: Relationship between location and dose levels. Bolded value is MTD in each order.

- $order = 5$: Alternating down-up diagonals

$$d_{11} \leq d_{12} \leq d_{21} \leq d_{31} \leq d_{22} \leq d_{13} \leq d_{23} \leq d_{32} \leq d_{41} \leq d_{42} \leq d_{33} \leq d_{43}$$

- $order = 6$: Alternating up-down diagonals

$$d_{11} \leq d_{21} \leq d_{12} \leq d_{13} \leq d_{22} \leq d_{31} \leq d_{41} \leq d_{32} \leq d_{23} \leq d_{33} \leq d_{42} \leq d_{43}$$

Now, we introduce two new complete orders:

- $order = 7$: Alternating down-up outer diagonals

$$d_{11} \leq d_{12} \leq d_{21} \leq d_{31} \leq d_{13} \leq d_{22} \leq d_{23} \leq d_{41} \leq d_{32} \leq d_{42} \leq d_{33} \leq d_{43}$$

- $order = 8$: Alternating up-down outer diagonals

$$d_{11} \leq d_{21} \leq d_{12} \leq d_{13} \leq d_{31} \leq d_{22} \leq d_{41} \leq d_{23} \leq d_{32} \leq d_{33} \leq d_{42} \leq d_{43}$$

In our simulation study, we check the performance of our new model with combination drugs, using the above listed 8 complete orders. These 8 complete orders

	True toxicity probabilities for 12 combination dose levels											
location / scenario	1	2	3	4	5	6	7	8	9	10	11	12
1	0.010	0.040	0.070	0.100	**0.200**	0.310	0.380	0.450	0.520	0.600	0.700	0.800
2	0.010	0.100	0.380	0.600	0.040	**0.200**	0.450	0.700	0.070	0.310	0.520	0.800
3	0.010	0.040	0.100	0.070	**0.200**	0.380	0.310	0.450	0.600	0.520	0.700	0.800
4	0.010	0.100	0.040	0.380	**0.200**	0.070	0.600	0.450	0.310	0.700	0.520	0.800
5	0.010	0.040	0.100	0.380	**0.200**	0.070	0.310	0.450	0.600	0.700	0.520	0.800
6	0.010	0.100	0.040	0.070	**0.200**	0.380	0.600	0.450	0.310	0.520	0.700	0.800
7	0.010	0.040	0.100	0.380	0.070	**0.200**	0.310	0.600	0.450	0.700	0.520	0.800
8	0.010	0.100	0.040	0.070	0.380	**0.200**	0.600	0.310	0.450	0.520	0.700	0.800

Table 5.3: True toxicity probabilities scenarios 1-8 for the combination dose levels, bolded value is target toxicity probability in each scenario.

	Skeleton probabilities for 12 combination dose levels											
location / order	1	2	3	4	5	6	7	8	9	10	11	12
1	0.004	0.020	0.050	0.110	**0.20**	0.310	0.420	0.530	0.630	0.720	0.780	0.840
2	0.0004	0.004	0.020	0.050	0.110	**0.200**	0.310	0.420	0.530	0.630	0.720	0.780
3	0.004	0.020	0.050	0.110	**0.200**	0.310	0.420	0.530	0.630	0.720	0.780	0.840
4	0.004	0.020	0.050	0.110	**0.200**	0.310	0.420	0.530	0.630	0.720	0.780	0.840
5	0.004	0.020	0.050	0.110	**0.200**	0.310	0.420	0.530	0.630	0.720	0.780	0.840
6	0.004	0.020	0.050	0.110	**0.200**	0.310	0.420	0.530	0.630	0.720	0.780	0.840
7	0.004	0.004	0.020	0.050	0.110	**0.200**	0.310	0.420	0.530	0.630	0.720	0.780
8	0.0004	0.004	0.020	0.050	0.110	**0.200**	0.310	0.420	0.530	0.630	0.720	0.780

Table 5.4: Skeletons for 8 orders, bolded value is the MTD in each skeleton.

are listed in Table 5.2 and their corresponding scenarios are given in Table 5.3. All scenarios are based on the same simulation data set in Table 5.1 We also introduce the initial guess of toxicity probabilities, called skeleton values for 8 scenarios, are provided in Table 5.4. The values are generated by "getprior(0.05,0.20,5,12)" when the MTD is located at 5 and "getprior(0.05,0.20,6,12)" when the MTD is located at 6.

5.4 Simulation study

To extend the power, logistic and hyperbolic tangent models to the case of drug combinations, we let η be a constant term in the logistic model, and α be positive. Let p_{ij} be the joint probability of combination dose level (i, j), $i = 1, 2, \cdots, K$ for Drug A and $j = 1, 2, \cdots, L$ for Drug B.

The power model at combination dose levels (i, j) is given by

$$P(\text{toxicity at combination dose levels } i, j) = p_{ij}^{exp(\alpha)}.$$

The logistic model at combination dose levels (i, j) is given by

$$\pi_{i,j} = \frac{exp(\eta + \alpha p_{ij})}{1 + exp(\eta + \alpha p_{ij})}.$$

The hyperbolic tangent model at combination dose levels (i, j) is given by

$$\pi_{i,j}(\alpha) = \left\{ \frac{(e^{2p_{ij}} - 1)/(e^{2p_{ij}} + 1) + 1}{2} \right\}^{\alpha}.$$

Our new model at combination dose levels (i, j) is given by

$$\pi_{ij} = \frac{2\Phi(\eta + \alpha p_{ij})}{1 + \Phi(\eta + \alpha p_{ij})}.$$

After treating n patients in the trial and observing at least one toxic outcome, if the data are analyzed under complete order s , then the likelihood function can be written as

$$L(\alpha|D_n) = \prod_{k=1}^{n} \{\pi_{ij}^s\}^{y_k} \{1 - \pi_{ij}^s\}^{1-y_k}.$$

Where k is the number of treated patients, $k = 1, \cdots, n$. π_{ij}^{s} is the joint toxicity probability at combination dose level (i, j) under complete order s, $i = 1, \cdots, 4$, $j = 1, 2, 3$ and $s = 1, \cdots, 8$ in our trial.

Suppose that α follows the prior distribution $f(\alpha)$, by the Bayes' Theorem, after treating n patients, the posterior mean toxicity probability at combination dose level (i, j) is estimated to be

$$\hat{\pi}(i,j) = \int \frac{2\Phi(\eta + \alpha p_{ij})}{1 + \Phi(\eta + \alpha p_{ij})} \frac{L(\alpha|D_n) f(\alpha)}{\int L(\alpha|D_n) f(\alpha) d\alpha} d\alpha.$$

In our simulation study, our goal is to compare the performance of our new model with other three classic models in the literature listed above with combination drugs under the performance criteria: BEARS. The simulation setting is as follows: (1) the sample size is $n = 30$; (2) the target DLT rate is $\theta = 0.2$; (3) there are 12 combination dose levels and 8 simulation scenarios based on possible complete orders are given in Table 5.3, together with 8 skeletons provided in Table 5.4. Each skeleton indicates one movement of escalation-deescalation. Here, we consider 8 possible orders of drug combinations. In our case, combination dose levels d_{22} is the MTD in every scenario. Because of different movements, the MTD is located at 5 in scenarios 1, 3, 4, 5 and 6, and the MTD is located at 6 in scenarios 2, 7 and 8, from Table 5.2; (4) the prior distribution for α is the $gamma(x, 0.5, 0.5)$ distribution in R; (5) each trial is simulated for 2,000 runs.

Benchmark means using sample mean toxicity probabilities to choose the MTD. In Table 5.5, we compare benchmark with the dose selection probability at

the MTD using sample mean toxicity probabilities. Similarly, in Table 5.6, we compare final percentage MTD selection using posterior mean toxicity probabilities.

- **Performance measure:** The dose with a sample (or posterior) mean toxicity probability nearest to θ is chosen as the MTD. The higher the proportion of correct selection the MTD, the better the design.

- **Results:** From Table 5.5 and Table 5.6, in scenarios 2, 6, 7 and 8, values of Benchmark are greater than 50%, which means more than 50% patients are truly treated at the MTD in our model. The benchmark is 61.8% in scenario 2, which is the highest value in all scenarios. In scenario 5, the benchmark of the MTD is lower than that of the higher combination dose levels d_{23} in our model and logistic model, which means we assign an overdose higher than MTD to treat patients based on these two models. However, the benchmark of the MTD is lower than that of the lower combination dose level d_{21} in the hyperbolic tangent model, which means we assign an ineffective dose lower than MTD to treat patients. Roughly speaking, the power model performs the best, and our model performs almost the second best in 8 scenarios. The reason may be that the more parameters in the model, the less robust the performance of design. Similarly, in scenario 6, the proportion of MTD selection is 74.8%, which is the highest in all scenarios. In scenarios 2, 6 and 7, values of the proportion of MTD selection are greater than 70% in

our model.

Efficiency checks the avoidance of assigning patients to ineffective doses.

- **Two performance measures:** E1 reports the proportion of simulation runs that assign patients to any dose lower than the MTD. The lesser the proportion E1, the better the design. E2 gives the mean and standard deviation of the percentage of patients assigned to any dose lower than the MTD. The lower the mean, the better the design.

- **Results:** From Table 5.7, Although, our model does not give the lowest values of E1 and E2, but relatively these values are around zero. Therefore, overall, the our model is efficient.

Accuracy depicts the assignment of patients to the MTD.

- **Three performance measures:** A1 measures the proportion of simulation runs that correctly select the MTD. The larger the proportion, the better the design. A2 reports the mean and standard deviation of the percentage of patients assigned to the MTD. The larger the mean, the better the design. A3 calculates the proportion of simulation runs with more than 50% of patients assigned to the MTD. The higher the proportion, the better the design.

- **Results:** From Table 5.7, in all scenarios, our model gives reasonably high values of A1, A2 and A3. values of A1 in scenarios 2, 6, 7 and 8 are higher

than 70%. In scenarios 1 and 3, values of A1 are near 50%. Values of A3 are higher than 50% in scenarios 2, 6, 7 and 8. Overall, Our model is accurate.

Reliability concerns the risk of severe overdosing.

- **Two performance measures:** R1 measures the proportion of simulation runs that allocate more than 50% of patients to any dose higher than the MTD. The lower the proportion, the better the design. R2 calculates the proportion of simulation runs with less than one-sixth of patients at the MTD. The lower the proportion, the better the design.

- **Results:** From Table 5.7, our model performs above the average in all scenarios. Particularly in scenarios 2, 6 and 7, our model shows less than 20% of total patients being assigned to doses higher than the MTD. Overall, our design is reliable.

Safety refers to the protection from overdosing.

- **Two performance measures:** S1 measures the proportion of simulation runs that result in any dose higher than the MTD. The lesser its value, the better the design. S2 reports the mean and standard deviation of the percentage of patients assigned to any dose higher than the MTD. The lower the mean, the better the design.

- **Results:** From Table 5.7, our model performs above the average in all

scenarios. In scenarios 2, 6, 7 and 8, our model shows lower risk of overdosing. Overall, our model design is safe.

Putting all criteria "BEARS" together, the performance of our design is overall stable. But the power model shows the best performance. Lasonos et al. (2016) indicate that the problem is caused by putting far more parameters into the model than those can be estimated. To further understand the advantage of our design, Figure 5.2 describes the comparison of estimated density curves of the proportion of patients truly selected the MTD based on 2,000 simulation runs. Our model performs well and stable.

5.5 Conclusion

In this chapter, we have extended our new model to the case of combination of two drugs. One particular difficulty for combination drug Phase I trials is to define a complete order among all drug combinations so that the escalation or de-escalation rule is clear. However, given several dose levels for each drug, the total dose combinations are complicated and there is no unique complete order available. So the goal is to introduce a one-dimensional order for a two-dimensional data structure.

Besides six complete orders in the literature, we introduce two new ones. We compare the performance of our new model and existing models with combination drugs. We have only included the interaction term. The marginal contributions

Scenario (MTD)		Benchmark (Skipping is not permitted)											
1	order / model	d_{11}	d_{12}	d_{13}	d_{21}	d_{22}	d_{23}	d_{31}	d_{32}	d_{33}	d_{41}	d_{42}	d_{43}
(2, 2)	I	0.150	0.001	0.000	0.054	0.353	0.103	0.255	0.059	0.014	0.010	0.001	0.000
	II	0.023	0.000	0.000	0.000	**0.657**	0.253	0.056	0.007	0.003	0.000	0.001	0.000
	III	0.138	0.002	0.000	0.002	0.231	0.100	0.321	0.129	0.051	0.020	0.006	0.000
	IV	0.048	0.018	0.054	0.217	0.368	0.195	0.084	0.015	0.001	0.000	0.000	0.000
2	order / model	d_{11}	d_{21}	d_{31}	d_{41}	d_{12}	d_{22}	d_{32}	d_{42}	d_{13}	d_{23}	d_{33}	d_{43}
(2,2)	I	0.092	0.003	0.000	0.000	0.102	0.618	0.093	0.001	0.012	0.072	0.007	0.000
	II	0.029	0.000	0.000	0.000	0.000	**0.924**	0.037	0.001	0.005	0.004	0.000	0.000
	III	0.121	0.002	0.000	0.000	0.008	0.379	0.053	0.001	0.130	0.250	0.056	0.000
	IV	0.180	0.292	0.183	0.018	0.028	0.253	0.043	0.002	0.001	0.000	0.000	0.000
3	order / model	d_{11}	d_{12}	d_{21}	d_{13}	d_{22}	d_{31}	d_{23}	d_{32}	d_{41}	d_{33}	d_{42}	d_{43}
(2,2)	I	0.071	0.000	0.001	0.050	0.402	0.074	0.330	0.057	0.000	0.013	0.002	0.000
	II	0.027	0.000	0.000	0.000	**0.728**	0.166	0.067	0.011	0.000	0.001	0.000	0.000
	III	0.088	0.000	0.000	0.002	0.228	0.060	0.446	0.122	0.019	0.029	0.006	0.000
	IV	0.061	0.024	0.058	0.167	0.457	0.117	0.095	0.021	0.000	0.000	0.000	0.000
4	order / model	d_{11}	d_{21}	d_{12}	d_{31}	d_{22}	d_{13}	d_{41}	d_{32}	d_{23}	d_{42}	d_{33}	d_{43}
(2,2)	I	0.444	0.002	0.002	0.053	0.206	0.156	0.020	0.034	0.082	0.001	0.000	0.000
	II	0.155	0.000	0.000	0.000	**0.404**	0.398	0.023	0.014	0.005	0.000	0.001	0.000
	III	0.337	0.001	0.000	0.002	0.058	0.229	0.036	0.123	0.180	0.017	0.017	0.000
	IV	0.267	0.094	0.109	0.146	0.099	0.245	0.034	0.004	0.002	0.000	0.000	0.000

Scenario (MTD)		Benchmark (Skipping is not permitted)											
5 (2,2)	order / model	d_{11}	d_{12}	d_{21}	d_{31}	d_{22}	d_{13}	d_{23}	d_{32}	d_{41}	d_{42}	d_{33}	d_{43}
	I	0.175	0.000	0.004	0.054	0.089	0.048	0.542	0.077	0.002	0.007	0.002	0.000
	II	0.069	0.000	0.000	0.000	**0.377**	0.314	0.222	0.015	0.003	0.000	0.000	0.000
	III	0.142	0.000	0.000	0.000	0.014	0.077	0.590	0.130	0.029	0.008	0.009	0.001
	IV	0.145	0.061	0.243	0.100	0.085	0.153	0.196	0.017	0.000	0.000	0.000	0.000
6 (2,2)	order / model	d_{11}	d_{21}	d_{12}	d_{13}	d_{22}	d_{31}	d_{41}	d_{32}	d_{23}	d_{33}	d_{42}	d_{43}
	I	0.121	0.001	0.000	0.049	0.594	0.141	0.006	0.021	0.059	0.007	0.001	0.000
	II	0.035	0.000	0.000	0.000	**0.774**	0.157	0.014	0.012	0.005	0.003	0.000	0.000
	III	0.149	0.000	0.000	0.006	0.407	0.143	0.028	0.098	0.138	0.031	0.000	0.000
	IV	0.080	0.013	0.023	0.203	0.514	0.152	0.013	0.000	0.001	0.001	0.000	0.000
7 (2,2)	order / model	d_{11}	d_{12}	d_{21}	d_{31}	d_{13}	d_{22}	d_{23}	d_{41}	d_{32}	d_{42}	d_{33}	d_{43}
	I	0.118	0.002	0.000	0.002	0.083	0.541	0.176	0.009	0.058	0.002	0.009	0.000
	II	0.025	0.000	0.000	0.000	0.000	**0.752**	0.213	0.007	0.003	0.000	0.000	0.000
	III	0.156	0.003	0.000	0.000	0.004	0.409	0.233	0.026	0.123	0.018	0.027	0.001
	IV	0.085	0.053	0.124	0.033	0.126	0.356	0.205	0.017	0.001	0.000	0.000	0.000
8 (2,2)	order / model	d_{11}	d_{21}	d_{12}	d_{13}	d_{31}	d_{22}	d_{41}	d_{23}	d_{32}	d_{33}	d_{42}	d_{43}
	I	0.138	0.001	0.000	0.002	0.083	0.538	0.024	0.177	0.028	0.006	0.003	0.000
	II	0.035	0.000	0.000	0.000	0.000	**0.923**	0.009	0.022	0.009	0.002	0.000	0.000
	III	0.061	0.001	0.000	0.000	0.006	0.359	0.015	0.394	0.121	0.032	0.011	0.000
	IV	0.148	0.028	0.047	0.238	0.185	0.279	0.017	0.054	0.004	0.000	0.000	0.000

Table 5.5: Simulated values of Benchmark from scenario 1 to scenario 8. The MTD in each scenario is located at d_{22}. Bold values are highest in each group.

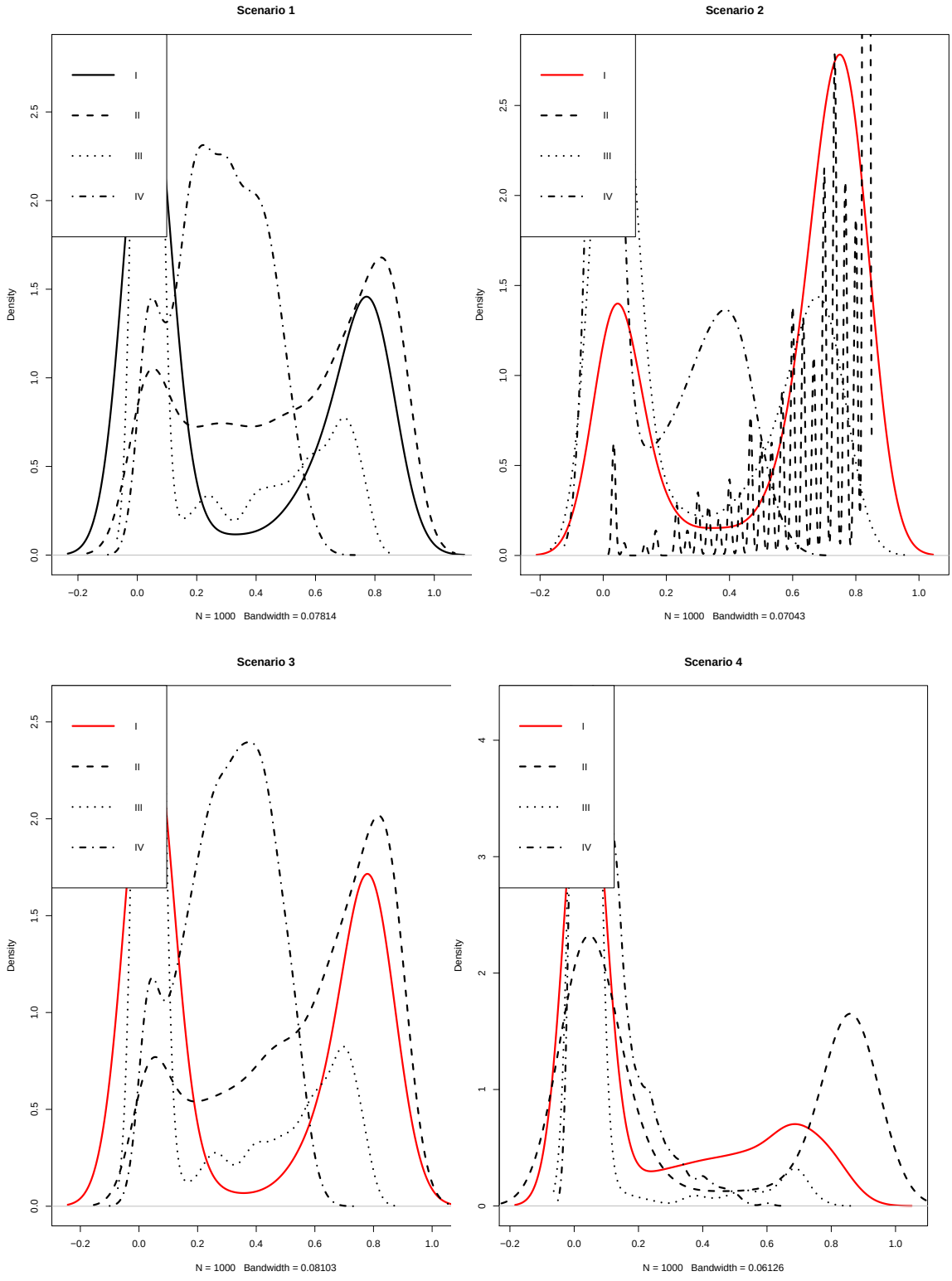
Scenario (MTD)		Proportion of MTD selection (Skipping is not permitted)											
1	order / model	d_{11}	d_{12}	d_{13}	d_{21}	d_{22}	d_{23}	d_{31}	d_{32}	d_{33}	d_{41}	d_{42}	d_{43}
(2, 2)	I	0.000	0.000	0.001	0.057	0.432	0.000	0.477	0.000	0.033	0.000	0.000	0.000
	II	0.000	0.000	0.000	0.000	**0.743**	0.233	0.023	0.001	0.000	0.000	0.000	0.000
	III	0.000	0.000	0.000	0.002	0.282	0.051	0.614	0.020	0.031	0.000	0.000	0.000
	IV	0.003	0.006	0.042	0.232	0.440	0.224	0.047	0.006	0.000	0.000	0.000	0.000
2	order / model	d_{11}	d_{21}	d_{31}	d_{41}	d_{12}	d_{22}	d_{32}	d_{42}	d_{13}	d_{23}	d_{33}	d_{43}
(2,2)	I	0.000	0.000	0.000	0.000	0.135	0.725	0.001	0.023	0.006	0.110	0.000	0.000
	II	0.000	0.000	0.000	0.000	0.000	**0.963**	0.029	0.000	0.004	0.004	0.000	0.000
	III	0.000	0.000	0.000	0.000	0.017	0.432	0.040	0.103	0.067	0.337	0.004	0.000
	IV	0.163	0.050	0.313	0.037	0.111	0.271	0.053	0.000	0.001	0.001	0.000	0.000
3	order / model	d_{11}	d_{12}	d_{21}	d_{13}	d_{22}	d_{31}	d_{23}	d_{32}	d_{41}	d_{33}	d_{42}	d_{43}
(2,2)	I	0.000	0.000	0.000	0.056	0.457	0.001	0.466	0.004	0.016	0.000	0.000	0.000
	II	0.000	0.000	0.000	0.000	**0.810**	0.165	0.022	0.003	0.000	0.000	0.000	0.000
	III	0.000	0.000	0.000	0.003	0.262	0.025	0.658	0.028	0.024	0.000	0.000	0.000
	IV	0.002	0.011	0.036	0.224	0.501	0.166	0.048	0.012	0.000	0.000	0.000	0.000
4	order / model	d_{11}	d_{21}	d_{12}	d_{31}	d_{22}	d_{13}	d_{41}	d_{32}	d_{23}	d_{42}	d_{33}	d_{43}
(2,2)	I	0.000	0.000	0.001	0.095	0.339	0.086	0.395	0.001	0.083	0.000	0.000	0.000
	II	0.000	0.000	0.000	0.000	**0.396**	0.494	0.103	0.002	0.004	0.001	0.000	0.000
	III	0.000	0.000	0.000	0.001	0.072	0.284	0.529	0.011	0.103	0.000	0.000	0.000
	IV	0.022	0.022	0.205	0.205	0.082	0.315	0.143	0.004	0.002	0.000	0.000	0.000

Scenario (MTD)		Proportion of MTD selection (Skipping is not permitted)											
5	order / model	d_{11}	d_{12}	d_{21}	d_{31}	d_{22}	d_{13}	d_{23}	d_{32}	d_{41}	d_{42}	d_{33}	d_{43}
(1,2)	I	0.000	0.000	0.001	0.107	0.122	0.001	0.747	0.006	0.016	0.000	0.000	0.000
	II	0.000	0.000	0.000	0.000	**0.361**	0.383	0.248	0.008	0.000	0.000	0.000	0.000
	III	0.000	0.000	0.000	0.000	0.016	0.044	0.893	0.027	0.020	0.000	0.000	0.000
	IV	0.030	0.043	0.222	0.152	0.057	0.232	0.243	0.020	0.001	0.000	0.000	0.000
6	order / model	d_{11}	d_{21}	d_{12}	d_{13}	d_{22}	d_{31}	d_{41}	d_{32}	d_{23}	d_{33}	d_{42}	d_{43}
(2,2)	I	0.000	0.000	0.000	0.055	0.748	0.007	0.121	0.001	0.068	0.000	0.000	0.000
	II	0.000	0.000	0.000	0.000	**0.869**	0.121	0.003	0.002	0.003	0.002	0.000	0.000
	III	0.000	0.000	0.000	0.008	0.517	0.068	0.306	0.010	0.090	0.001	0.000	0.000
	IV	0.002	0.004	0.022	0.247	0.536	0.181	0.006	0.001	0.000	0.001	0.000	0.000
7	order / model	d_{11}	d_{12}	d_{21}	d_{31}	d_{13}	d_{22}	d_{23}	d_{41}	d_{32}	d_{42}	d_{33}	d_{43}
(2,2)	I	0.000	0.000	0.002	0.001	0.091	0.708	0.018	0.172	0.002	0.006	0.000	0.000
	II	0.000	0.000	0.000	0.000	0.000	**0.835**	0.162	0.003	0.000	0.000	0.000	0.000
	III	0.000	0.000	0.000	0.000	0.003	0.548	0.148	0.292	0.006	0.003	0.000	0.000
	IV	0.022	0.021	0.076	0.079	0.185	0.392	0.210	0.015	0.000	0.000	0.000	0.000
8	order / model	d_{11}	d_{21}	d_{12}	d_{13}	d_{31}	d_{22}	d_{41}	d_{23}	d_{32}	d_{33}	d_{42}	d_{43}
(2,2)	I	0.000	0.000	0.001	0.002	0.172	0.561	0.000	0.247	0.000	0.017	0.000	0.000
	II	0.000	0.000	0.000	0.000	0.000	**0.958**	0.027	0.012	0.003	0.000	0.000	0.000
	III	0.000	0.000	0.000	0.000	0.013	0.399	0.007	0.551	0.018	0.012	0.000	0.000
	IV	0.023	0.026	0.052	0.346	0.288	0.192	0.065	0.005	0.003	0.000	0.000	0.000

Table 5.6: Simulated percentages of MTD selection from scenario 1 to scenario 8. The MTD in each scenario is located at d_{22}. Bold values are highest in each group.

Scenario		Efficiency		Accuracy			Reliability		Safety	
(MTD)	Model	E1	E2 (s.d.)	A1	A2 (s.d.)	A3	R1	R2	S1	S2 (s.d.)
1	I	0.058	0.045 (0.090)	0.432	0.330 (0.346)	0.397	0.470	0.558	0.510	0.070 (0.139)
(2,2)	II	0	0.033 (0)	0.743	0.507 (0.298)	0.539	0.213	0.187	0.257	0.051 (0.097)
	III	0.002	0.034 (0.150)	0.282	0.174 (0.242)	0.166	0.471	0.721	0.716	0.099 (0.116)
	IV	0.283	0.119 (0.082)	0.440	0.276 (0.146)	0.039	0.029	0.210	0.277	0.036 (0.071)
2	I	0.135	0.057 (0.117)	0.725	0.526 (0.312)	0.674	0.095	0.267	0.140	0.032 (0.088)
(2,2)	II	0	0.033 (0)	0.963	0.774 (0.137)	0.938	0.006	0.011	0.037	0.010 (0.047)
	III	0.017	0.035 (0.028)	0.432	0.278 (0.294)	0.333	0.084	0.561	0.551	0.091 (0.122)
	IV	0.674	0.160 (0.107)	0.271	0.146 (0.181)	0.018	0.001	0.607	0.055	0.009 (0.039)
3	I	0.056	0.045 (0.088)	0.457	0.363 (0.358)	0.442	0.441	0.530	0.487	0.065 (0.137)
(2,2)	II	0	0.033 (0)	0.810	0.562 (0.277)	0.623	0.125	0.135	0.190	0.044 (0.087)
	III	0.003	0.034 (0.016)	0.262	0.173 (0.245)	0.169	0.491	0.733	0.735	0.099 (0.114)
	IV	0.273	0.118 (0.080)	0.501	0.302 (0.148)	0.063	0.012	0.179	0.226	0.032 (0.067)
4	I	0.096	0.053 (0.120)	0.339	0.210 (0.271)	0.206	0.431	0.668	0.565	0.083 (0.137)
(2,2)	II	0	0.033(0)	0.396	0.384 (0.383)	0.400	0.498	0.483	0.604	0.069 (0.125)
	III	0.001	0.033 (0.007)	0.072	0.072 (0.143)	0.049	0.447	0.926	0.927	0.113 (0.131)
	IV	0.454	0.149 (0.110)	0.082	0.094 (0.100)	0.002	0.131	0.799	0.464	0.044 (0.095)
5	I	0.108	0.056 (0.128)	0.122	0.115 (0.221)	0.107	0.699	0.874	0.770	0.095 (0.125)
(2,2)	II	0	0.033 (0)	0.361	0.063 (0.380)	0.373	0.381	0.506	0.639	0.072 (0.126)
	III	0	0.033 (0)	0.016	0.043 (0.067)	0.011	0.678	0.980	0.984	0.118 (0.098)
	IV	0.447	0.149 (0.109)	0.057	0.081 (0.093)	0.002	0.053	0.826	0.496	0.046 (0.095)
6	I	0.055	0.044 (0.086)	0.748	0.523 (0.303)	0.651	0.184	0.239	0.197	0.043 (0.107)
(2,2)	II	0	0.033 (0)	0.869	0.604 (0.248)	0.681	0.095	0.083	0.131	0.038 (0.079)
	III	0.008	0.035 (0.024)	0.517	0.305 (0.288)	0.330	0.200	0.495	0.475	0.080 (0.113)
	IV	0.275	0.122 (0.080)	0.536	0.328 (0.130)	0.070	0.010	0.095	0.189	0.026 (0.055)
7	I	0.094	0.047 (0.094)	0.708	0.473 (0.300)	0.580	0.157	0.276	0.198	0.049 (0.106)
(2,2)	II	0	0.033 (0)	0.835	0.606 (0.270)	0.687	0.137	0.118	0.165	0.040 (0.097)
	III	0.003	0.034 (0.015)	0.548	0.278 (0.262)	0.280	0.209	0.490	0.449	0.092 (0.111)
	IV	0.383	0.120 (0.084)	0.392	0.229 (0.154)	0.013	0.014	0.338	0.225	0.029 (0.069)
8	I	0.175	0.059 (0.125)	0.561	0.446 (0.352)	0.562	0.233	0.410	0.264	0.043 (0.123)
(2,2)	II	0	0.033 (0)	0.958	0.713 (0.170)	0.894	0.007	0.027	0.042	0.020 (0.051)
	III	0.013	0.035 (0.027)	0.399	0.261 (0.289)	0.307	0.394	0.601	0.588	0.094 (0.116)
	IV	0.735	0.148 (0.089)	0.192	0.172 (0.166)	0.038	0.001	0.548	0.073	0.015 (0.046)

Table 5.7: Simulated values of EARS, not allowing skipping.


Scenario 1
Density
I
II
III
IV
N = 1000 Bandwidth = 0.07814
Scenario 2
Density
I
II
III
IV
N = 1000 Bandwidth = 0.07043
Scenario 3
Density
I
II
III
IV
N = 1000 Bandwidth = 0.08103
Scenario 4
Density
I
II
III
IV
N = 1000 Bandwidth = 0.06126

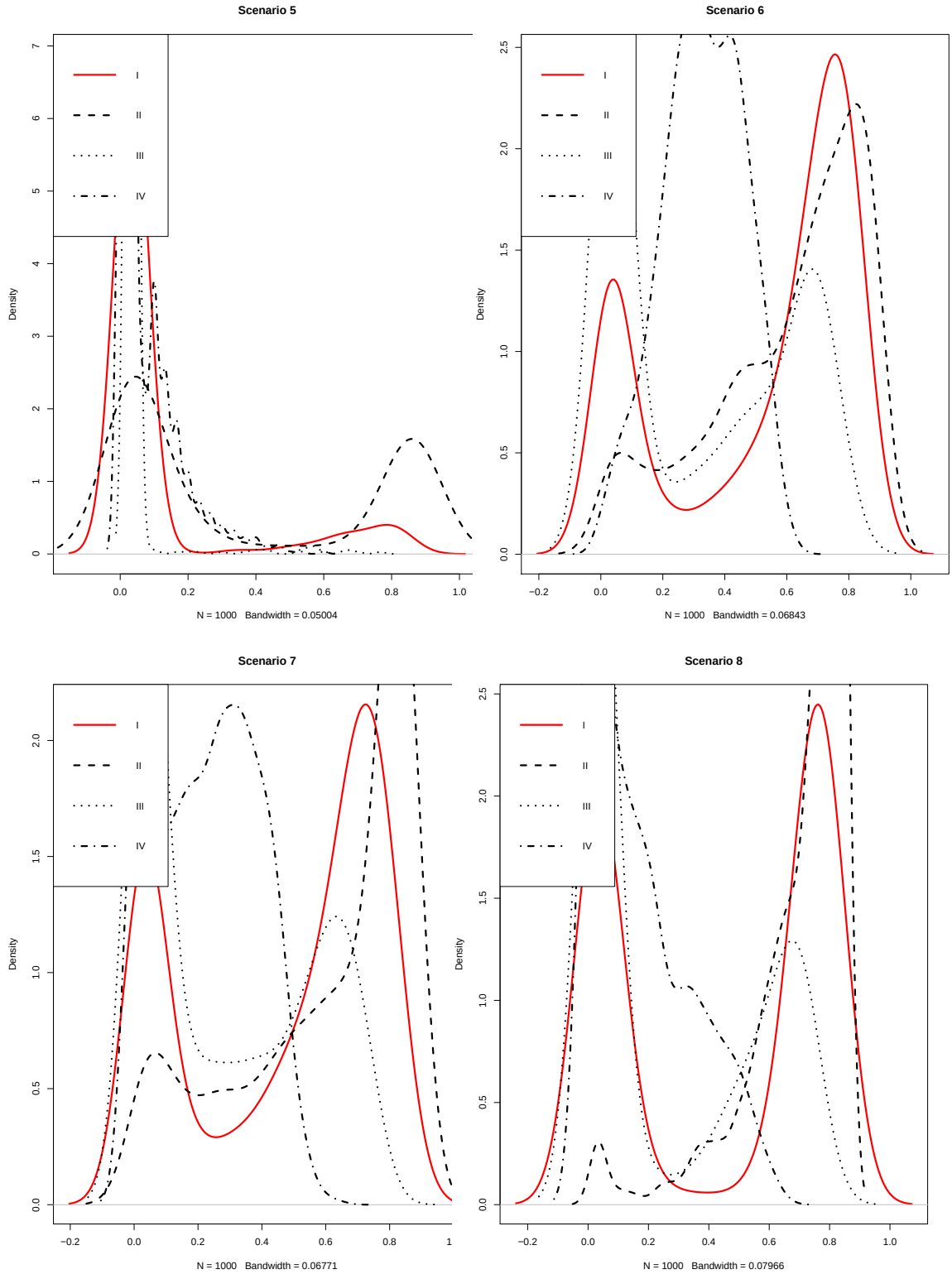


Figure 5.2: Comparison of density estimation of MTD allocation, Scenario 1 to 8, no skipping

by each drug is not considered because the data structure dose not provide the marginal distributions. This is an area that is worth further investigation.

Chapter 6

Statistical Inference

6.1 Introduction

In previous chapters, we have focused on introducing new designs of the continual reassessment method (CRM) and evaluating their statistical performance. Our new designs are based on the cumulative distribution function of the normal distribution, and our new model involves one unknown parameter α in the basic model with one drug.

Although our new designs identify the maximum tolerated dose (MTD) at the end of trial, it is only a point estimation and from statistical point of view, it is important to develop statistical procedures for the confidence interval and hypothesis testing for the population MTD.

The fundamental goal of these statistical procedures is to establish the sampling distribution of the statistic for the unknown parameter α. Since Phase I clinical trials often involve small sample sizes, exact methods based on small

sample sizes are often difficult to derive. Most statistical procedures are using asymptotic statistical method.

This is the key point of this current chapter. In this chapter, using tools of asymptotic statistical inference, we derive the asymptotic distribution of the sample statistic a for estimating the population parameter α. This asymptotic distribution forms the foundation for deriving the confidence interval and hypothesis testing procedure for the unknown parameter α.

6.2 Maximum Likelihood Estimation

Harville (1977) indicates that model parameters can be estimated by generalized least squares of α and restricted maximum likelihood estimation. Wooldridge (2010) proposes the asymptotic normality and asymptotic variance estimation under random sampling. To estimate unknown parameters by maximum likelihood estimators (MLE), we start with the likelihood function. The likelihood function is given by

$$L(\alpha) = f(y_i|d_i;\alpha) = \prod_{i=1}^{n} \left\{ \frac{2\Phi(\beta+\alpha d_i)}{1+\Phi(\beta+\alpha d_i)} \right\}^{y_i} \left\{ 1 - \frac{2\Phi(\beta+\alpha d_i)}{1+\Phi(\beta+\alpha d_i)} \right\}^{1-y_i},$$

where $y_i = 0, 1$.

The log likelihood function can be written as

$$\begin{aligned} & l(\alpha) \\ = & \sum_{i=1}^{n} log\left\{f(y_i|d_i;\alpha)\right\} \\ = & \sum_{i=1}^{n} y_i log\left\{\frac{2\Phi(\beta+\alpha d_i)}{1+\Phi(\beta+\alpha d_i)}\right\} + (1-y_i) log\left\{1-\frac{2\Phi(\beta+\alpha d_i)}{1+\Phi(\beta+\alpha d_i)}\right\}. \end{aligned}$$

The range of $\frac{2\Phi(\cdot)}{1+\Phi(\cdot)}$ is between 0 and 1, which ensures that the log-likelihood function is well defined. The MLE of α, denoted by $\hat{\alpha}$, maximizes the log-likelihood function and is given by

$$\hat{\alpha} = argmax\ log(\alpha) \quad = \quad argmax \sum_{i=1}^{n} log f(y_i|d_i;\alpha).$$

As usual, the MLE of α does not have a closed form expression but can be obtained by the Newton-Raphson algorithm. To calculate the MLE, we introduce the score function

$$\begin{aligned} & s(\alpha) \\ = & \sum_{i=1}^{n} s_i(\alpha) \\ = & \sum_{i=1}^{n} \frac{dlog\left\{f(y_i|d_i;\alpha)\right\}}{d\alpha} \\ = & \sum_{i=1}^{n} \frac{\left\{y_i+\Phi(\beta+\alpha d_i)y_i - 2\Phi(\beta+\alpha d_i)\right\}\phi(\beta+\alpha d_i)d_i}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))(1+\Phi(\beta+\alpha d_i))} \\ = & \sum_{i=1}^{n} \frac{\phi(\beta+\alpha d_i)d_i Z_i}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))}, \end{aligned}$$

where $Z_i = y_i - \pi_i = y_i - \frac{2\Phi(\beta+\alpha d_i)}{1+\Phi(\beta+\alpha d_i)}$ and y_i is the toxicity outcome associated with d_i for $i = 1, 2, \cdots, n$. Now y_i takes value 1 with probability π_i and 0 with

probability $1-\pi_i$. Since $E(Z_i|d_i)=0$, it follows that

$$\begin{aligned}
& E(s_i(\alpha)|d_i) \\
= & \frac{\{1+\Phi(\beta+\alpha d_i)-2\Phi(\beta+\alpha d_i)\}\,\phi(\beta+\alpha d_i)d_i}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))(1+\Phi(\beta+\alpha d_i))}\left\{\frac{2\Phi(\beta+\alpha d_i)}{1+\Phi(\beta+\alpha d_i)}\right\} \\
+ & \frac{\{-2\Phi(\beta+\alpha d_i)\}\,\phi(\beta+\alpha d_i)d_i}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))(1+\Phi(\beta+\alpha d_i))}\left\{1-\frac{2\Phi(\beta+\alpha d_i)}{1+\Phi(\beta+\alpha d_i)}\right\} \\
= & 0.
\end{aligned}$$

Assuming that $l(\alpha)$ is a twice continuously differentiable function, the asymptotic variance of $\hat{\alpha}$ is given by

$$\begin{aligned}
H_n(\alpha) &= \frac{ds(\alpha)}{d\alpha} \\
&= \sum_{i=1}^{n}\frac{d^2 log_i(\alpha)}{d\alpha^2} \\
&= -\sum_{i=1}^{n}\frac{\phi(\beta+\alpha d_i)^2 d_i^2}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))} \\
&\quad +Z_i\frac{d}{d\alpha}\left\{\frac{\phi(\beta+\alpha d_i)d_i}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))}\right\}.
\end{aligned}$$

The expected value of the asymptotic variance conditional on d_i is shown to be

$$-E\left\{\frac{1}{n}H_n(\alpha_0)|d_i\right\}=\frac{\{\phi(\beta+\alpha_0 d_i)\}^2 d_i^2}{\Phi(\beta+\alpha_0 d_i)(1-\Phi(\beta+\alpha_0 d_i))},$$

because

$$\begin{aligned}
& E\left\{Z_i\frac{d}{d\alpha}\left\{\frac{\phi(\beta+\alpha d_i)d_i}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))}\right\}|d_i\right\} \\
= & E(Z_i|d_i)\frac{d}{d\alpha}\left\{\frac{\phi(\beta+\alpha d_i)d_i}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))}\right\}=0.
\end{aligned}$$

Let α_0 be the true value of the parameter α. By the first order Taylor expansion,

$$S_n(\hat{\alpha}) \approx S_n(\alpha_0) + \frac{dS_n(\bar{\alpha})}{d\alpha}(\hat{\alpha} - \alpha_0),$$

where $\bar{\alpha}$ lies between $\hat{\alpha}$ and α_0, hence, $\bar{\alpha}$ converges to α_0 in probability.

We multiply by $\frac{1}{\sqrt{n}}$ on both sides, then

$$\begin{aligned} \frac{1}{\sqrt{n}} S_n(\hat{\alpha}) &\approx \frac{1}{\sqrt{n}} S_n(\alpha_0) + \frac{1}{n} \frac{dS_n(\bar{\alpha})}{d\alpha} \sqrt{n}(\hat{\alpha} - \alpha_0) \\ &= \frac{1}{\sqrt{n}} S_n(\alpha_0) + \frac{1}{n} H_n(\bar{\alpha}) \sqrt{n}(\hat{\alpha} - \alpha_0). \end{aligned}$$

If $\bar{\alpha} \xrightarrow{p} \alpha_0$, then $\frac{1}{n} H_n(\bar{\alpha}) \xrightarrow{p} H(\alpha_0)$ and $\hat{\alpha}$ is the solution to $S_n(\hat{\alpha}) = 0$, so

$$\frac{1}{\sqrt{n}} S_n(\alpha_0) + H(\alpha_0) \sqrt{n}(\hat{\alpha} - \alpha_0) + o_p(1) = 0.$$

It follows that

$$\sqrt{n}(\hat{\alpha} - \alpha_0) = -H(\alpha_0)^{-1} \frac{1}{\sqrt{n}} S_n(\alpha_0) + o_p(1).$$

We define

$$\begin{aligned} A &= -E\left\{H_i(\alpha)\right\} \\ &= -E\left\{\frac{d^2 log_i(\alpha)}{d\alpha^2}\right\} \\ &= -E\left\{\frac{\phi(\beta + \alpha d_i)^2 d_i^2}{\Phi(\beta + \alpha d_i)(1 - \Phi(\beta + \alpha d_i))}\right\}, \end{aligned}$$

and

$$\begin{aligned} B &= Var\left\{\frac{dlog_i(\alpha)}{d\alpha}\right\} \\ &= E\left\{\frac{d^2log_i(\alpha)}{d\alpha^2}\right\} - \left\{E\left\{\frac{dlog_i(\alpha)}{d\alpha}\right\}\right\}^2 \\ &= E\left\{\frac{d^2log_i(\alpha)}{d\alpha^2}\right\} \\ &= E\left\{s_i(\alpha)^2\right\} \\ &= E\left\{\frac{Z_i^2\phi(\beta+\alpha d_i)^2 d_i^2}{\Phi(\beta+\alpha d_i)^2(1-\Phi(\beta+\alpha d_i))^2}\right\}. \end{aligned}$$

Then

$$\begin{aligned} E\left\{\frac{dlog_i(\alpha)}{d\alpha}\middle| d_i\right\} &= \int \frac{dlog_i(\alpha)}{d\alpha} f(y_i|d_i;\alpha)dy \\ &= \int \frac{1}{f(y_i|d_i;\alpha)} \frac{df(y_i|d_i;\alpha)}{d\alpha} f(y_i|d_i;\alpha)dy \\ &= \int \frac{df(y_i|d_i;\alpha)}{d\alpha} dy \\ &= \frac{d}{d\alpha}\int f(y_i|d_i;\alpha)dy \\ &= 0. \end{aligned}$$

By the law of total expection,

$$E\left\{\frac{dlog_i(\alpha)}{d\alpha}\right\} = E\left\{E\left\{\frac{dlog_i(\alpha)}{d\alpha}\middle| d_i\right\}\right\} = 0,$$

so

$$E(s_i(\alpha|d_i)) = \int s_i(\alpha) f(y_i|d_i;\alpha)dy = 0.$$

From the above identity, we take derivatives on both sides, we have

$$\begin{aligned}
\frac{d}{d\alpha}\int s_i(\alpha)f(y_i|d_i;\alpha)dy &= \int \frac{d}{d\alpha}s_i(\alpha)f(y_i|d_i;\alpha)dy \\
&= \int \left\{\frac{ds_i(\alpha)}{d\alpha}f(y_i|d_i;\alpha) + s_i(\alpha)\frac{df(y_i|d_i;\alpha)}{d\alpha}\right\}dy \\
&= \int \frac{ds_i(\alpha)}{d\alpha}f(y_i|d_i;\alpha)dy + \int s_i(\alpha)\frac{df(y_i|d_i;\alpha)}{d\alpha}dy \\
&= \int H_i(\alpha)f(y_i|d_i;\alpha)dy + \int s_i(\alpha)\frac{dlogf(y_i|d_i;\alpha)}{d\alpha}f(y_i|d_i;\alpha)dy \\
&= E\left\{H_i(\alpha)|d_i\right\} + E\left\{s_i(\alpha)^2\middle|d_i\right\} \\
&= 0.
\end{aligned}$$

Therefore,

$$-E\left\{H_i(\alpha)|d_i\right\} = E\left\{s_i(\alpha)^2\middle|d_i\right\},$$

and

$$-E\left\{H_i(\alpha)\right\} = E\left\{s_i(\alpha)^2\right\}, which\ implies\ A = B.$$

For sufficiently large n, $\sqrt{n}(\hat{\alpha}-\alpha_0) \xrightarrow{d} N(0, A^{-1}BA^{-1})$. For $A = B$, $\sqrt{n}(\hat{\alpha}-\alpha_0) \xrightarrow{d} N(0, A^{-1})$, where $A = -E\left\{s_i(\alpha_0)\right\}$.

The asymptotic variance, denoted by $AVar(\hat{\alpha})$, of $\hat{\alpha}$ is estimated as

$$AVar(\hat{\alpha}) = \left\{\sum_{i=1}^{n}\frac{\left\{\phi(\beta+\alpha d_i)\right\}^2 d_i^2}{\Phi(\beta+\alpha d_i)(1-\Phi(\beta+\alpha d_i))}\right\}^{-1}.$$

Therefore, $\hat{\alpha}$ asymptotically follows a normal distribution with mean α_0 and variance $\frac{1}{n}\hat{AVar}(\alpha)^{-1}$, denoted by $\hat{\alpha} \overset{Asy.}{\sim} N(\alpha_0, \frac{1}{n}\hat{AVar}(\alpha)^{-1})$.

The (asymptotic) confidence interval of α is given by

$$\hat{\alpha} \pm Z^* \sqrt{\left\{ \sum_{i=1}^{n} \frac{\{\phi(\beta + \hat{\alpha} d_i)\}^2 d_i^2}{\Phi(\beta + \hat{\alpha} d_i)(1 - \Phi(\beta + \hat{\alpha} d_i))} \right\}^{-1}},$$

where Z^* is the upper critical value in the standard normal distribution.

On the other hand, the asymptotic statistic for testing the null hypothesis $H_0 : \alpha = \alpha_0$ versus the alternative hypothesis $H_a : \alpha \neq \alpha_0$ ($\alpha > \alpha_0$ *or* $\alpha < \alpha_0$) is given by

$$Z = \frac{\hat{\alpha} - \alpha_0}{\sqrt{\left\{ \sum_{i=1}^{n} \frac{\{\phi(\beta + \alpha_0 d_i)\}^2 d_i^2}{\Phi(\beta + \alpha_0 d_i)(1 - \Phi(\beta + \alpha_0 d_i))} \right\}^{-1}}},$$

which follows the standard normal distribution.

6.3 Numerical Illustration

We use the Newton-Raphson method to find the MLE of parameter α in the one drug model discussed in Chapter 2. For example, the second row in Table 6.1 shows that in scenario 1, after 348 iterations, the Newton-Raphson procedure stopped, and $\hat{\alpha}$ converged to 2.999996. The 95% asymptotic confidence interval of α is calculated to be $(1.07, 4.93)$. This suggests that we may use $\alpha = 3$ in future studies.

6.4 Conclusion

To follow up with our new proposed designs of the CRM, we develop the foundation for statistical inference of our new designs. The center of this development is

Scenario	n.stop	$\hat{\alpha}$	C.I.
1	348	2.999996	$(1.07, 4.93)$
2	209	2.999994	$(0.00, 7.91)$
3	203	2.999994	$(1.09, 4.91)$
4	197	2.999994	$(0.00, 7.50)$
5	133	2.999991	$(0.00, 8.62)$
6	16	2.999996	$(0.00, 15.40)$

Table 6.1: values of $\hat{\alpha}$ in scenarios 1 to 6

to derive the asymptotic distribution of the sample statistic a for the unknown parameter α, and this is achieved in this chapter using techniques of asymptotic statistical methods. Finally, we derive the asymptotic confidence interval and hypothesis testing procedure for the unknown parameter α.

Chapter 7

Summary

In my PhD Thesis, I have introduced some new parametric designs of Phase I clinical trials and evaluated their performance.

The new designs use the Continual Reassessment Method (CRM) but are based on new parametric models depicting the toxicity probability function of the dose level. The new function shows similar analytic properties of the existing parametric functions in the literature that have been used with CRM. However, our new function represents not just a single model but instead a class of models. This allows the possibility of choosing the most appropriate model with this class.

We also introduce some new performance measures and summarize our comparison criteria as BEARS: Benchmark, Efficiency, Accuracy, Reliability and Safety. Such criteria allow us to systematically compare and evaluate the performance of our new models and existing models.

In summary, simulation results show that our new designs perform relatively

good in comparison with currently available designs in the literature, based on the criteria BEARS. Our new designs are potentially useful in the practice of clinical trials, however, there may be still practical issues to resolve.

At the end, we obtain the asymptotic distribution of the unknown parameter in our new model and construct the foundation for deriving the confidence interval and hypothesis testing procedure for the unknown parameter.

In future, we will consider the overdose and ineffective dose controls in Phase I. We will choose other cumulative distribution functions to replace the normal distribution in our proposed model, and we can even use a semi-parametric model to replace the normal distribution in Phase I. We will also consider the late-onset toxicity in Phase I and late-onset efficacy in Phase II, or combined seamless PhaseI/II clinical trials. Finally, we can investage theoretical properities of our new models.

Bibliography

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.

Bailey, S., B., N., Laird, G. and Branson, M. (2009). A Bayesian case study in oncology phase i combination dose-finding using logistic regression with covariates. *Journal of Biopharmaceutical Statistics*, **19** 469–484.

Chen, Z., Tighiouart, A. 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.

Cheung, Y. (2011). *Dose finding by the continual reassessment method*. CRC Press, New York, USA.

Cheung, Y. (2014). Simple benchmark for complex dose finding studies. *Biometrics*, **70** 389–397.

Clertant, M. and O'Quigley, J. (2017). Semiparametric dose finding methods. *Journal of the Royal Statistical Society: Series B*, **79** 1487–1508.

Clertant, M. and O'Quigley, J. (2019). Semiparametric dose finding methods: special cases. *Applied Statistics*, **68(2)** 271–288.

Dancey, J. and Chen, H. (2006). Strategies for optimizing combinations of molecularly targeted anticancer agents. *Nature Reviews. Drug Discovery*, **5** 649–659.

Diniz, M., Li, Q. and Tighiouart, M. (2017). Dose finding for drug combination in early cancer phase i trials using conditional continual reassessment method. *Journal of Biometrics & Biostatistics*, **8(6)** 1000381.

Durham, S., Flournoy, N. and Rosenberger, W. (1997). A random walk rule for phase i clinical trials. *Biometrics*, **53** 745–760.

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

Gezmu, M. and Flournoy, N. (2006). Group up-and-down designs for dose-finding. *Journal of Statistical Planning and Inference*, **136** 1749–1764.

Hamberg, P., Ratain, M., Lesaffre, E. and Verweij, J. (2010). Dose-escalation models for combination phase i trials in oncology. *European Journal of Cancer*, **46** 2870–2878.

Harrington, J., Wheeler, G., Sweeting, M., Mander, A. and Jodrell, D. (2013). Adaptive designs for dual-agent phase i dose-escalation studies. *Nature Reviews. Clinical Oncology*, **10** 277–288.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72** 320–338.

Hirakawa, A., Wages, N., Sato, H. and Matsui, S. (2014). A comparative study of adaptive dose-finding designs for phase i oncology trials of combination therapies. *Statistics in Medicine*, **34** 3194–3213.

Hoeting, J., Madigan, D., Raftery, A. E. and Volinsky, C. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, **14** 382–401.

Iasonos, A. and O'Quigley, J. (2014). Adaptive dose-finding studies: a review of model-guided phase i clinical trials. *Journal of Clinical Oncology*, **23** 2505–2511.

Iasonos, A., Wilton, A., Riedel, E., Seshan, V. and Spriggs, D. (2008). A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in phase i dose-finding studies. *Clinical Trials*, **5** 465–477.

Ivanova, A., Flournoy, N. and Chung, Y. K. (2007). Cumulative cohort design for dose finding. *Journal of Statistical Planning and Inference*, **137** 2316–2327.

Kramar, A., Lebecq, A. and Candalh, E. (1999). Continual reassessment

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

Lasonos, A., Wages, N., Conaway, M., Cheung, K., Yuan, Y. and O'Quigley, J. (2016). Dimension of model parameter space and operating characteristics in adaptive dose-finding studies. *Statistics in Medicine*, **35** 3760–3775.

Le Tourneau, C., Lee, J. and Siu, L. (2009). Dose escalation methods in phase I cancer clinical trials. *Journal of the National Cancer Institute*, **101** 708–720.

Lee, S. and Cheung, Y. (2009). Model calibration in the continual reassessment method. *Clinical Trials*, **6** 227–238.

Leung, D. and Wang, Y. (2002). An extension of the continual reassessment method using decision theory. *Statistics in Medicine*, **21** 51–63.

Lin, Y. and Shih, W. (2001). Statistical properties of the traditional alogrithm-based designs for phase I cancer clinical trials. *Biostatistics*, **2** 203–215.

Marchenko, O., Fedorov, V., Lee, J., Nolan, C. and Pinheiro, J. (2013). Adaptive clinical trials: overview of early-phase designs and challenges. *Therapeutic Innovation & Regulatory Science*, **48** 20–30.

Marusyk, A., Almendro, V. and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? *Nature Reviews. Cancer*, **12** 323–334.

MASTER, J. S. and KÖBERLE, B. (2003). Curing metastatic cancer: lessons from testicular germ-cell tumours. *Nature Reviews. Cancer*, **3** 517–525.

MDANDERSON (Accessed 2019-05-13). Clinical trials at MD Anderson: Is a clinical trial the right treatment choice for you? URL https://www.mdanderson.org/patient-and-cancer-information/cancer-information/clinical-trials/clinicaltrials-1111.pdf.

NEUENSCHWANDER, B., BRANSON, M. and GSPONER, T. (2008). Critical aspects of the bayesian approach to phase I cancer trials. *Statistics in Medicine*, **27** 2420–2439.

O'QUIGLEY, J. (1992). Estimating the probability of toxicity at the recommended dose following a phase I clinical trial in cancer. *Biometrics*, **48** 853–863.

O'QUIGLEY, J. (2002). Continual reassessment designs with early termination. *Biostatistics*, **3** 87–99.

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.

O'QUIGLEY, J. and CONAWAY, M. (2010). Continual reassessment and related dose-finding designs. *Statistical Science*, **25** 202–216.

O'QUIGLEY, J., PEPE, M. and FISHER, L. (1990). Continual reassessment

method: a practical design for phase I clinical trials in cancer. *Biometrics*, **46** 33–48.

O'Quigley, J. and Reiner, E. (1998). A stopping rule for the continual reassessment method. *Biometrika*, **85** 741–748.

O'Quigley, J. and Shen, L. (1996). Continual reassessment method: a likelihood approach. *Biometrics*, **52** 673–684.

Paller, C., Bradbury, P., Ivy, S., Seymour, L., LoRusso, P., Baker, L., Rubinstein, L., Huang, E., Collyar, D., Groshen, S., Reeves, S., Ellis, L., Sargent, D., Rosner, G., LeBlanc, M. and Ratain, M. (2014). Design of phase I combination trials: recommendations of the clinical trial design task force of the nci investigational drug steering committee. *Clinical Cancer Research*, **20** 4210–4217.

Pan, H. and Yuan, Y. (2017). A default method to specify skeletons for bayesian model averaging continual reassessment method for phase I clinical trials. *Statistics in Medicine*, **36** 266–279.

Potter, D. (2006). Phase I studies of chemotherapeutic agents in cancer patients: a review of the designs. *Journal of Biopharmaceutical Statistics*, **16** 579–604.

Raftery, A. E., Madigan, D. and Hoeting, J. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, **92** 179–191.

Ratain, M., Mick, R., Schilsky, R. and Siegler, M. (1993). Statistical and ethical issues in the design and conduct of phase i and ii clinical trials of new anticancer agents. *Journal of the National Cancer Institute*, **85** 1637–1643.

Riviere, M., Duboisb, F. and Zohar, S. (2015). Competing designs for drug combination in phase i dose-finding clinical trials. *Statistics in Medicine*, **34** 1–12.

Rosenberger, W. and Haines, L. (2002). Competing designs for phase i clinical trials: a review. *Statistics in Medicine*, **21** 2757–2770.

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

Simon, R., Freidlin, B., Rubinstein, L., Arbuck, S., Collins, J. and Christian, M. (1997). Accelerated titration designs for phase i clinical trials in oncology. *Journal of the National Cancer Institute*, **89** 1138–1147.

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

Sverdlov, O., Wong, W. and Ryeznik, Y. (2014). Adaptive clinical trial designs for phase i cancer studies. *Statistics Surveys*, **8** 2–44.

Sweeting, M. and Pander, A. (2012). Escalation strategies for combination therapy phase i trials. *Pharmaceutical Statistics*, **12** 217–224.

Tighiouart, M. and Rogatko, A. (2010). Dose finding with escalation with overdose control (ewoc) in cancer clinical trials. *Statistical Science*, **25** 217–226.

Tighiouart, M., Rogatko, A. and Babb, J. (2005). Flexible Bayesian methods for cancer phase i clinical trials. dose escalation with overdose control. *Statistics in Medicine*, **24** 2183–2196.

Wages, N. (2015). Comments on 'competing designs for drug combination in phase i dose-finding clinical trials' by m-k. Riviere, f. Dubois and s. Zohar. *Statistics in Medicine*, **34** 18–22.

Wages, N. and Conaway, M. (2013). Specifications of a continual reassessment method design for phase i trials of combined drugs. *Pharmaceutical Statistics*, **12** 217–224.

Wages, N., Conaway, M. and O'Quigley, J. (2011). Dose-finding design for multi-drug combinations. *Clinical Trials*, **8** 380–389.

Wages, N., Ivanova, A. and Marchenko, O. (2016). Practocal designs for phase i combination studies in oncology. *Journal of Biopharmaceutical Statistics*, **26** 150–166.

Weber, J., Levit, L., Adamson, P., Bruinooge, S., Burris, H., Carducci, M., Dicker, A., Gönen, M., Keefe, S., Postow, M., Thompson, M., Waterhouse, D., Weiner, S. and Schuchter, L. (2015). American society

of clinical oncology policy statement update: the critical role of phase I trials in cancer research and treatment. *Journal of Clinical Oncology*, **33** 278–284.

Wheeler, G., Sweeting, M. and Mander, A. (2017). Toxicity-dependent feasibility bounds for the escalation with overdose control approach in phase I cancer trials. *Statistics in Medicine*, **36** 2499–2513.

Whitehead, J. and Brunier, H. (1995). Bayesian decision procedures for dose determining experiments. *Statistics in Medicine*, **14** 885–893.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data*. Massachusetts Institute of Technology.

Yin, G., and Lin, R. (2015). Comments on 'competing designs for drug combination in phase I dose-finding clinical trials' by M-K. Riviere, F. Dubois and S. Zohar. *Statistics in Medicine*, **34** 13–17.

Yin, G. (2012). *Clinical trial design: Bayesian and frequentist adaptive methods*. Wiley.

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

Yin, G. and Yuan, Y. (2009b). Bayesian model averaging continual reassrssment method in phase I clinical trials. *Journal of the American Statistical Association*, **104** 954–968.

Yuan, Y. and Yin, G. (2011). Bayesian phase I/II adaptively randomized oncology trials with combined drugs. *Annals of Applied Statistics*, **5(2A)** 924–942.

Zacks, S., Rogatko, A. and Babb, J. (1998). Optimal Bayesian-feasible dose escalation for cancer phase I trials. *Statistics and Probability Letters*, **38** 215–220.

Zhang, W., Wang, X. and Yang, P. (2019). A new design of the continual reassessment method. *Communications in Statistics - Simulation and Computation*.

Zhou, H., Yuan, Y. and Nie, L. (2018). Accuracy, safety, and reliability of novel phase I trial designs. *Clinical Cancer Research*, **24** 4357–4364.