

The CRM (Continual Reassessment Method) Designs in the
Presence of Population Heterogeneity


Jianfen Shu
Quzhou, China

B.S. Peking University, 1997
M.S. Georgetown University, 2000
M.S. University of Minnesota, 2002

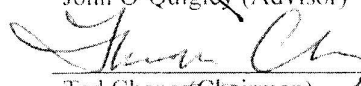
A Dissertation Presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Statistics
University of Virginia


August 2012



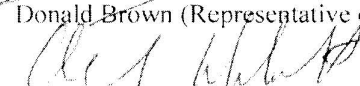
John O'Quigley (Advisor)



Ted Chang (Chairman)



Donald Brown (Representative of Dean)



Robert Abbott



Guofen Yan

Abstract

This thesis first studies the parameterization issue in the CRM. The nature of the CRM design causes patients to be accumulated on a single dose [1], which leads to insufficient data to estimate two parameters. The two-parameter model is too flexible and has potentially erratic behaviors, which cause more early patients receiving toxic and untested doses. The performance of the two-parameter CRM is highly influenced by prior specification which is often expressed as pseudo-data. Our study shows the two-parameter CRM with the commonly used pseudo-data performs poorly for very safe treatments, compared with the two-stage one-parameter CRM. It is more reassuring to use one stage of the trial to collect the prior instead of specifying the prior or using imaginary patients.

The main focus of this thesis is to investigate the extensions of CRM to two-group or multiple-group studies where patients are from two or more different populations. Bridging studies, in which there are two distinct groups, such as adults and children, and where the adult study precedes the study in children, are a special case of the heterogeneity problem. In many two-group studies, it is known that one group of patients is less sensitive to treatment and the true MTD is higher compared to the other group. In this research, we focus on the CRM shift model. In this model, the MTD for one group is shifted either up or down by a certain number of dose levels depending on the known information about the group difference. The operating characteristic of this CRM model is demonstrated by simulated examples and performance is evaluated together with other CRM approaches and the optimal design. The shift model can be applied to situations with more than 2 groups. Stopping rules and method robustness are explored. Proof of asymptotical consistency with realistic conditions is provided as well.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. John O'Quigley. Without his constant support and guidance, this dissertation could not have been completed.

I would like to thank my other committee members: Professor Ted Chang, Professor Robert Abbott, Professor Guofen Yan and Professor Donald Brown. I am also grateful to Professor Jeff Holt. Without his suggestion and encouragement, I would not have become a PhD student in the Department of Statistics.

I would like to thank my family in China and my friends for their continued support and encouragement. Most importantly, I would like to thank my husband, Mir, for providing his support and sharing my sadness, frustration and happiness. I also feel grateful to have two dearest kids, Zevash and Sheerin. Your smiles and sweet words have made my challenging journey more enjoyable.

Jianfen Shu

June 20, 2012

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Phase I Oncology Studies	1
1.2 Designs for Phase I Oncology Studies	4
1.2.1 Algorithm-Based Designs	4
1.2.2 Model-Based Designs	10
1.3 Topics to Be Studied	13
1.3.1 One-group CRM Parameterization	14
1.3.2 Heterogeneity and Bridging studies	14
1.4 Outline of the Dissertation	18
2 Parameterization of the One-group CRM Model	19
2.1 Details of the Standard One-group CRM model	19
2.2 One-parameter versus Two-parameter	23
2.2.1 Lack of Stability of the Two-parameter CRM	25
2.2.2 Performance of the Two-parameter CRM with Pseudo-data for Shallow Dose-toxicity Curves	28

3	Two-group CRM Shift Model	31
3.1	Scheme I: CRM Shift model	31
3.1.1	A Motivating Example	31
3.1.2	CRM Shift Model Detail	32
3.2	Scheme II: Two-parameter CRM Model	38
3.3	Scheme III: Two Groups, Two Separate Trials	40
3.4	Scheme IV: Ignore the Group Difference and Pool the Two Groups	40
4	Two-group CRM Model Performance	41
4.1	Comparison of Four Schemes for Two-group Studies	41
4.2	Using Prior Information for the Levels Shifted	44
4.3	Comparison of the CRM Shift Model and the Optimal Design . . .	44
4.3.1	Incomplete Information	45
4.3.2	Complete Information	46
4.3.3	Near Optimality of the CRM Shift Model Performance . . .	47
5	Three-group CRM Shift Model	48
5.1	Equal Probability of Shift and Non-shift	53
5.2	Unequal Probability of Shift and Non-shift	54
6	Stopping Rules for the CRM Shift Model	55
7	Robustness of the CRM Shift Model	59
7.1	The Effect of Classification	60
7.2	The Effect of Misclassification	63
8	Consistency of the CRM Shift Model	65
8.1	The Set-up	66

8.2	Specifying Conditions	70
8.3	Theorem	73
8.4	Proof of Theorem	73
8.4.1	Consistent Model Selection	73
8.4.2	Consistent Dose Selection	81
9	Summary and Remarks	86
A	Tables	91
B	Figures	107
	Bibliography	119

Chapter 1

Introduction

1.1 Phase I Oncology Studies

Drug development is a continuous process through which the knowledge of efficacy and safety of an experimental drug is gradually accumulated. In the investigation of new drugs and treatments, phase I dose-finding trials aim at identifying a safe and efficient drug administration in humans, before proceeding to phase II and phase III randomized clinical trials focused on treatment efficacy. It is essential to have sufficient information of one or two doses of the drug that can be suitable for safety and efficacy in order to proceed with confirmatory trials. For common ailments, phase I studies most often use healthy volunteers, anticipating only mild side effects which should be entirely reversible. Phase I oncology clinical trials are usually performed on subjects for whom all currently available therapies have failed. While the new experimental treatments provide hope to these patients, the potential efficacy is almost always accompanied by life-threatening toxicities from the treatments. These trials are typically small, single-arm, open-

label, sequential studies. A Phase I oncology trial represents the first test of the treatment in patients, although there may be experience with closely related treatments. The classic oncology Phase I trial is only intended to determine the dose and characterize the side effects, not to show whether the treatment is effective, but, of course, if it turns out the treatment actually is effective, patients can still benefit. The outcome of a phase I cancer trial is binary dose-limiting toxicity (DLT) which is often defined as grade 3 unexpected treatment related toxicity or any toxicity greater than grade 3. The goal of phase I dose-finding trials is to locate the so-called maximum tolerated dose (MTD), the dose closest to the pre-specified toxicity rate [2]. The target rate typically ranges from 0.20 to 0.35. A phase I trial design involves many components, including starting dose, dose specification, dose escalation, cohort size, specification of DLT, target toxicity rate, definition of the MTD, and patient selection.

Phase I cancer clinical trial designs begin with choosing a starting dose based on animal studies. The starting dose is usually a certain percentage of low toxic dose in animals or based on information from individual drugs if combinations of several drugs are used and their drug information is available. The subsequent dose levels follow a modified Fibonacci scheme [3].

Mostly often dose-finding studies are first in human trials and little is known about the new drug. The safety of the participants is of primary concern. Ethical concerns have significantly influenced the sample size and the design of these studies. Typically sample sizes of dose-finding studies are small, commonly about one or two dozen. The limited sample size and the ethical issues are the two most critical problems currently facing phase I oncology studies.

Some rather special designs are used in dose-finding. The designs have long

been governed by the ethical constraint of minimizing subjects treated at toxic doses. Those designs are sequential designs instead of randomized designs. Randomly assigning patients between several dose levels will put some patients on low doses that are known to have suboptimal efficacy and other patients on high doses that are too toxic. Therefore, in addition to finding the right MTD, another goal of the phase I oncology trials (the therapeutic aim) is to maximize the chance that the dose that a patient receives should not only have the potential therapeutic value but also be safe. The process of dose escalation is governed by a fundamental conflict. There is a need to balance the concern for patient safety when being treated with an unknown agent, as reflected in careful dose escalation and the desire to treat at doses that will be close to the recommended phase II dose, thus increasing the likelihood of benefit. The ideal is to accumulate treatment assignment on or around the MTD. A good study design should escalate the dose fast enough so that patients would not receive ineffective low doses but at the same time avoid putting patients on doses with unacceptable toxicity.

Many designs have been proposed during the last two decades to address this question. These designs can be grouped into two broad classes of designs for phase I oncology studies: Algorithm-Based Approaches and Model-Based Approaches. The philosophical and statistical issues are different between the two classes of designs. Algorithm-based designs define the MTD as a statistic computed from data while the model-based consider the MTD as a parameter of a monotonic dose-response curve to be estimated from data. All designs are based on the monotonic dose-toxicity relationship which holds for oncology trials. A typical dose-toxicity relationship is shown in Figure B.1. In the plot, if the target rate is 0.20, then the target dose is the *4th* pre-specified dose.

1.2 Designs for Phase I Oncology Studies

1.2.1 Algorithm-Based Designs

Traditional Designs

The well-known 3 + 3 design is a special case of the traditional $M + N$ designs without dose de-escalation (Figure B.2). First 3 patients are treated at the first dose. If there is no DLT among the 3 patients on the i th dose, the next dose will be escalated to the $i + 1$ level. If there is more than 1 DLT, dose level $i - 1$ will be declared as the MTD. If there is 1 DLT, an additional 3 patients will be recruited to the same i th dose. If there is no DLT among these 3 new patients, the dose will be escalated, otherwise, dose level $i - 1$ will be declared as the MTD. The algorithm for the $M + N$ designs with dose-escalation is similar to those without dose de-escalation. The difference is that the $(i - 1)$ th dose will be the next dose instead of being declared as the MTD for the situations of more than 1/3 or 1/6 DLT on the i th dose. With dose de-escalation, more patients will be permitted to be treated at a lower dose when excessive DLT incidences occur at the current dose level. In these traditional $M + N$ designs, the current dose has no chance at all to be declared as the MTD. The previous lower dose level is always declared as the MTD. Several variants have been proposed, including allowing the investigators to declare a current dose level as the MTD and using an intermediate dose as the starting dose [4].

Storer was among the first to study characteristics of the traditional designs from a statistician's standpoint. Storer also proposed several variants on the method including a two-stage procedure [5] [6]. In the two-stage design, a fast

initial stage with a cohort size of 1 is pursued until a toxicity is found. Then assignment proceeds in groups of three until a predetermined sample size is reached. They are designed to estimate the MTD as a 33rd percentile. Reiner *et al.* studied the operating characteristics of the traditional design. The authors conclude that the probability of stopping at an incorrect level is higher than generally believed [7]. The 3 + 3 design is designed to estimate the MTD as 33rd percentile. The traditional design lacks the flexibility of estimating other percentiles. The MTD from the traditional design is underestimated. The escalation is slow and can spend a lot of patients at low-toxicity (ie, potentially sub-therapeutic) doses while few patients actually receive doses at or near the recommended dose selected for phase II trials. This problem is especially obvious when the number of doses is large and the MTD is a high dose level. The main advantage of the traditional 3 + 3 design is that it is simple to implement and safe [8].

Accelerated Titration Designs

Intra-patient dose escalation is not commonly practiced in oncology phase I trials. Sheiner and coworkers suggested the use of intra-patient dose escalation to maximize the possibility of individual patients receiving efficacious doses and to increase the accuracy of the analysis of the phase I data. They also suggested using dose-response models in the analysis of phase I trials [9] [10] [11]. Analysis models that account for inter-patient and intra-patient variability separately will accommodate the possibility of cumulative toxicity and allowed for the construction of dose-toxicity curves for both the sensitive and typical patients. Simon *et al.* developed a family of “accelerated titration designs” and proposed the use of an accompanying dose-toxicity model [12]. The goal of accelerated titration

is to conduct rapid intra-patient dose escalation in order to reduce the number of under-treated patients [in the trials themselves] and provide a substantial increase in the information obtained. If a first dose does not induce toxicity, a patient may be escalated to a higher subsequent dose. After the MTD is determined, a final “confirmatory” cohort is treated at a fixed dose. Since it gives an MTD for each patient, it provides an idea about how MTDs vary between patients. A summary of the accelerated titration designs and intra-patient dose-escalation options is displayed in Table A.1. The main distinguishing properties of these designs are: 1) a quick initial escalation phase, 2) intra-patient dose escalation and 3) the ability to analyze trial results using a dose-toxicity model that incorporates parameters for inter-patient and intra-patient variation in toxicity and cumulative toxicity.

Jordan *et al.* studied intra-patient escalation of carboplatin in ovarian cancer patients and found “The median MTD documented here using intra-patient dose escalation ... is remarkably similar to that derived from conventional phase I studies.” In other words, accelerated titration seems to work [13]. Simon *et al.* [12] evaluated the designs by simulation comparing the accelerated titration designs and the traditional 3 + 3 design. The accelerated titration designs had a major reduction in the number of under-treated patients while it had very small increases in the average number of patients experiencing DLT or unacceptable toxicity. The average numbers of patients were also significantly reduced in the accelerated titration designs. The accelerated titration designs with intra-patient dose-escalation provides patients accrued early in the trial a full opportunity to be treated at a therapeutic dose. They are especially more effective in situations where inter-patient variability in susceptibility to toxicity is large.

One major concern of using the accelerated titration designs with intra-patient dose-escalation is about cumulative toxicity. The potential ill effects from intra-patient dose-escalation warrants caution from the clinical trial team.

Designs Based on Random Walk Rules

Traditional methods rely on up-and-down rules, but are not designed around any percentile of interest. Random walk rules generalize the up-and-down approach by providing a unifying theory for targeting any quantile of interest [14] [15]. The most common up-and-down procedures are 1st-order Markovian procedure where the rule for allocating the next dose relies only on the current treatment and its recent outcomes. The most appropriate random walk rule for phase I trials is perhaps the following rule [15]:

If patient $j - 1$ has been assigned to dose level d_i , assign patient j as follows. If patient $j - 1$ experienced a toxic response, then assign patient j to level $d_i - 1$. If patient $j - 1$ had no toxic response, then flip a biased coin with probability of heads $b \in [0, 0.5]$. If it lands heads up, assign patient j to level $d_i + 1$. If it lands heads down, assign patient j to level d_i . When the current dose is the lowest or the highest one and the rule would cause a treatment to be outside of the dose range, the current treatment is repeated instead.

By the law of large numbers of regular Markov chains, the proportions of patients (π) assigned to each dose converge to the stationary treatment distribution. Gezmu and Flournoy [16] proved that π is unimodal and the mode is near the target dose when the probability of head b for the biased coin is equal to $(1 - 2\theta)(1 - \theta)^{-1}$, where θ is the target rate.

Other Algorithm-Based Designs

Pharmacologically guided dose escalation (PGDE) method is another variation of the traditional 3 + 3 design that has not been widely used in clinical practice [17]. The PGDE method assumes that DLT can be predicted by plasma drug concentration and this relationship in humans can be accurately reflected by animal models. This method has two stages. In the first stage, a prespecified plasma exposure is extrapolated from preclinical data and the real-time pharmacokinetic (PK) data is obtained for each patient to determine plasma exposure. As long as the prespecified plasma exposure is not reached or no DLT occurs, dose escalation proceeds with one patient per dose level and typically at 100% dose increment. When the target exposure is reached or DLT occurs, stage 2 with a traditional 3 + 3 design kicks in. The PGDE method has not been widely adopted due to practical obstacles. It is difficult to obtain real-time PK results and is often problematic to extrapolating preclinical PK data to human studies with different treatment schedules. O'Quigley *et al.* extended the CRM to model the pharmacokinetic parameters in dose-finding studies [18]. Simulations showed the method performed satisfactorily.

The “rolling six” design (RSD) was originally proposed as a way to shorten the timeline of pediatric phase I trials by reducing the number of times a study is suspended to accrual [19]. The rolling six design allows for accrual of two to six patients concurrently onto a dose level based on the number of patients currently enrolled and evaluable, the number experiencing dose-limiting toxicity (DLT), and the number still at risk of developing a DLT. This design is for trials with prior information about the dose range to be evaluated. In the simulated scenarios, the rolling six design outperformed the 3 + 3 method in key metrics such as the time

to complete the study. On the other hand, on average, three more patients were enrolled onto the rolling six trials. There were no differences in the rates of DLTs. Hartford *et al.* [20] pointed out that the statistical significance of the findings is not clear, and either larger simulated analysis or actual clinical studies will be needed to determine whether the RSD indeed shortens the study time. Zhao *et al.* [21] compared the RSD with the time-to-event CRM (TITE-CRM). They demonstrated that the TITE-CRM identified the MTD more accurately than the RSD and did not increase the probability of exposing patients to toxic doses. They concluded that the TITE-CRM provides better accuracy than the RSD while providing a safe design for pediatric oncology phase I trials.

Summary of Algorithm-Based Designs

The main advantages of rule-based methods are that they are easy to implement and do not require special software. Generally, the trial can be implemented without the involvement of a statistician. However, as mentioned before, their performance is not guaranteed. Also these designs may be inefficient in establishing the dose that meets a specific target toxicity level. In addition, information only from the current dose level but not all available information is used in the decision of dose allocation for future patients as well as the definition of the recommended dose for phase II trials. The recommended MTD is then selected from the pre-specified dose levels depending on which one best fits the definition of acceptable toxicity set a priori. The rule-based methods are still popular and have been used in establishing safe recommended doses for phase II trials during the past several decades for anticancer agents that were eventually used worldwide in clinical practice. However, its success has been debated [8].

1.2.2 Model-Based Designs

The Continual Reassessment Method

As mentioned above, the ethical concerns dictate that we give the best possible treatment to the patients participating in the phase I studies. Therefore the statistical goals of any proposed model should be to estimate the MTD (d_0) consistently and efficiently, and during the course of the studies, concentrate as many experiments around d_0 as possible. O'Quigley, Pepe and Fisher [1] proposed the continual reassessment method (CRM) to meet the above requirements. The basic idea is to use all data from the study itself and all information available prior to the study to fit a statistical model and to estimate the model parameter. Then the estimated toxicity rates for each dose can be calculated. Next patient or cohort will receive the current best dose, the dose with toxicity closest to the target rate. The model details are shown in Section 2.1, Details of the Standard One-group CRM Model. The CRM model can be in a Bayesian framework with a prior for the parameter or in a pure likelihood framework with a 2-stage design. The CRM has gained popularity since its proposal. A large number of modifications or extensions have been developed. The CRM method should be considered as a class of methods rather than a single method, the members of the class depending on arbitrary quantities determined by the investigator such as: the spacing between the doses, the starting dose, cohort size, the form of the model, the initial dose escalation scheme in two stage designs or the prior selection for the Bayesian formulations.

The typical behavior of the CRM designs is displayed in Figure B.3. Patients enter sequentially to receive the treatment recommended by the model until a

fixed number of subjects have been treated or until a pre-specified stopping rule is met. Simulation studies [1] [22] [23] [24] [25] [8] have shown the operating characteristics of the CRM to be great, in terms of accuracy of final recommendation while simultaneously reducing the numbers of over-treated and under-treated patients.

Bayesian Decision Approaches

All other model-based designs proposed so far can be considered as derivatives of the standard CRM designs. The basic ideas are the same as the CRM. The Bayesian decision approaches introduced by Whitehead and Brunier [26] is essentially the CRM under the Bayesian framework. Whitehead *et al.* suggested using the two-parameter logistic model instead of the one-parameter model in the Bayesian decision approaches. The research about whether a two-parameter model should be used is in Section 2.2. We generally do not support using a two-parameter model for the CRM.

Escalation with Over-Dose Control

The approach of dose-escalation with overdose control (EWOC) was proposed by Babb, Rogatko and Zacks [27]. The main idea of the EWOC is the same as the CRM except that it uses a two-parameter location-scale family to model the dose-toxicity relationship and implements an over-dose control. The design calculates the probability that the dose selected for the next patient is higher than the target dose. Denote this probability as α . The dose d selected by EWOC meets the criteria $P(d > \theta) = \alpha$ when dose is a continuum. θ is the unknown target

dose. The selected dose is the largest dose with the criteria $P(d > \theta) \leq \alpha$ when dose is pre-specified categorical. Low values of α make the escalation cautious and high values cause larger steps. One should consider gradually increasing α during the course of the trial. In the beginning of a trial, there is a higher level of uncertainty about the target dose. Consequently, at the onset of a phase I trial, the probability of exceeding the target dose is typically set to a low value (e.g., 0.2) in order to minimize the possibility of overdosing patients. As the trial progresses, uncertainty about the target dose declines and one should increase α value. Because of overdose control in EWOC, more patients will be treated at low, possibly no or low efficacy doses as a price of minimizing over-dosing. The dose escalation to the MTD in EWOC is not as quick as in the CRM.

Time-to-Event CRM

Most dose-limiting toxicities are acute events that occur soon after delivery of study drugs. However, if an agent causes late-onset or cumulative toxicities, an undesirably large number of patients may be treated at toxic doses before any toxicity is observed. It is impractical to mandate that phase I trial designs have an extended assessment period to monitor for late or cumulative toxicity, as this would result in suspension of the trial and the dose-finding trial can be impractically long. Cheung and Chappell [28] developed a modification of the continual reassessment method known as the time to event continual reassessment method (TITE-CRM) that incorporates the time to DLT for each patient. They redefined the toxicity rate at dose d_i to be the probability of observing toxicity at d_i during a time period of length T after initiation of therapy. The time length is defined as the time that has elapsed from the moment when the i th patient entered the trial

to the time $(n + 1)$ th patient enters the trial. It was claimed that for treatments with late onset toxicity, the TITE-CRM is more efficient than the traditional 3+3 design or the continual reassessment method for determining the MTD and leads to shorter trial durations [29]. It is not true that the CRM design requires the complete follow-up of the current patient (or group) before admitting a new patient (or group). The new patient (or group) can be assigned to the current recommended dose and start treatment right away. The claim that TITE-CRM leads to shorter trial duration cannot be warranted.

Summary of Model-Based Designs

Model-based designs use all toxicity information accumulated during the trial and estimate the probability of toxicity based on some parametric models. Simulations have shown that model-based methods achieve good estimations of the target probability of dose-limiting toxicity at the recommended dose for phase II trials without treating too many patients at suboptimal doses. Implementing model-based method is less straightforward and requires statistical expertise and computing software on site to perform model fitting in real time.

1.3 Topics to Be Studied

This dissertation focuses on the well-accepted CRM design proposed on O'Quigley, Pepe, and Fisher [1]. Two main topics are covered: parameterization of the one-group CRM model and the CRM shift model in the presence of patient heterogeneity, with the latter as the main focus.

1.3.1 One-group CRM Parameterization

In the original one-group CRM work, the authors studied both one- and two-parameter models with the one-parameter model as the main focus. The true mechanism generating the observations could deviate from the one-parameter working model overall, however, close to the target toxicity rate, the true situation and the working model coincide fairly well [1]. The primary goal of phase I trials is to locate the MTD. Therefore, only the accuracy of the toxicity estimate of the dose closest to the target rate matters most. The two-parameter CRM is the logistic model and its advantage is of being straightforward and familiar to most people. However, the model-based designs have a special data pattern: data is clustered around a single dose. This phenomenon causes the two-parameter model not to be an ideal one for the CRM. This dissertation will cover the advantages and disadvantages of different parameterizations.

1.3.2 Heterogeneity and Bridging studies

In the case of two or more distinct groups, there can be several ways to proceed. Of course, it is always possible to simply carry out single studies in each of the groups separately. Such an approach makes the least amount of assumptions of the data, but can be inefficient and costly to carry out in practice. This is particularly so in pediatric studies where, often, there are very few children available to study compared to the adult equivalent population. If some basic assumptions can be made then it is possible to share information between the studies enabling greater use to be made of the available observations. In the case of bridging studies, the idea is to bridge information between, say, the adult and

pediatric populations. There is often an added component here in that the adult study may have already been completed before undertaking the study in children and so the observations are not gathered simultaneously on the two populations. We will see below that the CRM shift model, essentially the same model in both populations, but shifted upwards one or more doses, or downwards one or more doses, can enable very real gains in precision and performance.

As in other types of clinical trials we are essentially looking for an average effect and finding an average MTD. Even though very often heterogeneity can be unobserved, it is reasonably common to run into substantial patient heterogeneity unless we set up extremely tight inclusion criteria. In phase I trials it is not uncommon that patients be categorized into two or more groups, although making distinctions among more than 2 groups may not be practically feasible giving the typically small number of patients available. Furthermore, we often encounter the situations where a major goal for the studies is to find out the correct MTDs for different groups of patients. A common example, arising in cancer studies, concerns the degree of previous treatment which is often related to the patient's tolerability to a new treatment regimen. Another example occurs where the concomitant treatment is different, with one group receiving a cytotoxic agent while the other one receiving hormone therapy. In many situations, even though we do not know the true dose-toxicity relation, we may know something about the difference between the groups. For example, we may know with confidence that patients receiving hormone therapy can tolerate the treatment under study better than those receiving a cytotoxic agent. If the effect of heterogeneity is not strong, it can just be ignored. However when the effect is strong, ignoring the difference between the groups will lead to an average MTD for both groups. This MTD will

be too toxic for the less-tolerant group while being of less than optimal efficacy for the more tolerant group. A straightforward solution is to run two separate studies for each group. However, this approach does not utilize the common information between the groups and is statistically less efficient. For very small samples, we need a better solution to carry out a more refined analysis and to identify the most appropriate level for both groups. It is generally difficult to apply any algorithm-based design, such as the up-and-down scheme, in the presence of population heterogeneity unless two separate trials are carried out. The model-based CRM design provides better solutions to these situations.

An extension from the CRM model has been developed by O'Quigley, Shen and Gamst [30], specifically for this problem. In the article, the authors developed a method to allow the two groups of patients to be treated in the same study and to identify the appropriate MTDs for the two groups simultaneously. There is no assumption made about the order of sensitivity towards the treatments between the two groups. In O'Quigley and Paoletti [31], the authors introduced a two-parameter CRM model for ordered groups, incorporating the known information about the group differences, in particular the direction and magnitude of the difference, by introducing and controlling the second parameter in addition to the one parameter in the standard CRM method. Although no publication has been done, the idea of adding a new parameter to the model to represent group difference can be applied to the model based designs of EWOC and TITE-CRM.

In this thesis, we will explore another CRM extension for two-group or multi-group studies, the CRM shift model, which was introduced by O'Quigley [32]. The CRM shift model can be used with or without information about order of sensitivity towards the new treatments. In this model, a single parameter is used

to model the dose-response relationship for one group and restriction is set up for the difference between the two groups. The difference takes values from a small finite set. If the first group has identified dose level d_0 as the best recommended dose, then the other group will be recommended with either the same level or some level, one, two or more, steps away from d_0 . The study investigators can provide constraints on the direction and magnitude of the difference (how many doses one group deviates from the other one) between the two groups. The idea of the shift model is unique to the under-parameterized CRM and cannot practically be applied to other model based designs such as EWOC and TITE-CRM. Section 3.1 demonstrates the model details and its implementation. The idea is that such a specification will make better use of information common to both groups and make it easier for clinicians to understand the design. Also the information from clinicians about group difference can be easily incorporated into the model. Of course, we also want the method to behave well if it turns out that there is little or no shared information or if the difference between the two groups turns out to be small. We want to address the following questions:

1. For the situations where it is possible to carry out two separate studies, how does the CRM shift model compare to the standard one-sample CRM applied to the two groups separately?
2. Besides being easier for clinicians to understand the study design, how does the CRM shift model compare to the established two-parameter two-group CRM design?
3. If the difference between the two groups turns out to be small (in particular the bridging problem), how does the CRM shift model compare to the one-sample CRM carried out on the two groups combined?

4. Can the CRM shift model be of use when subgroup size imbalance is so strong that separate trials are not plausible? Again, this is an important consideration, in the context of bridging, for example in pediatric studies where there is likely to be many less children in the study than adults.
5. How does the CRM shift model behave when the sample sizes for both groups increase infinitely?

Besides addressing the above questions, we will compare the performance of the CRM shift model with the optimal design. The unique stopping rules for the CRM shift model and the robustness of this model are explored as well.

1.4 Outline of the Dissertation

The rest of the dissertation is organized as follows: Chapter 2 focuses on the one-group CRM while Chapters 3, 4, 5, 6, 7, and 8 study the CRM shift model. Specifically, Chapter 2 reviews the details of the standard one-group CRM and compares two different parameterizations: one-parameter versus two-parameter. Chapter 3 displays the details of the two-group CRM shift model and its implementation and Chapter 4 compares performance of different schemes handling population heterogeneity. Chapter 5 is the application of the CRM shift model to the 3-group situations. Chapter 6 compares different stopping rules that are unique to the multi-group CRM models and Chapter 7 examines model robustness including the effect of classification and the effect of misclassification. The asymptotic consistency of the CRM shift model is covered in Chapter 8 with a detailed proof included. Chapter 9 summarizes the dissertation.

Chapter 2

Parameterization of the One-group CRM Model

This chapter covers the model details for the standard one-group CRM model and compares two different parameterizations: one-parameter versus two-parameter.

2.1 Details of the Standard One-group CRM model

The CRM for the standard one-group studies uses a one-parameter under-parameterized model to identify the MTD and maximally concentrate patients around the MTD. Assuming we have k candidate doses ordered in toxicity rate, we denote $R(d_i)$ as the toxicity rate for dose d_i . There is no need for equally spaced doses because the CRM does not rely on the scale in which the doses are expressed. The dose for the j th patient X_j takes values $x_j \in \{d_1, \dots, d_k\}; j = (1, \dots, n)$. Let Y_j be the dose limiting toxicity (1=DLT, 0=No DLT) for patient j . We

model $R(x_j)$, the probability of toxicity, at $X_j = x_j$; $x_j \in \{d_1, \dots, d_k\}$ by

$$R(x_j) = Pr(Y_j = 1|X_j = x_j) = E(Y_j|x_j) = \psi(x_j, a)$$

for some one-parameter models $\psi(x_j, a)$. This parametric model should be flexible enough to approximate the underlying true dose-toxicity relationship in the neighborhood of the target toxicity rate, θ . Choosing a particular model $\psi(x, a)$ is largely a matter of personal taste and would not matter as long as $\psi(x, a)$ satisfying the following conditions [33].

1. For each a , functions $\psi(\cdot, a)$ is strictly increasing.
2. Function $\psi(x, \cdot)$ is continuous and is strictly monotone in a in the same direction for all x .
3. For each $0 < t < 1$ and each x , the function

$$s(t, x, a) = t \frac{\psi'}{\psi}(x, a) + (1 - t) \frac{-\psi'}{1 - \psi}(x, a)$$

is continuous and is strictly monotone in a . $\psi'(x, a)$ is the partial derivative of ψ with respect to a .

4. The parameter a belongs to a finite interval $[A, B]$.

There are an infinite number of models satisfying the above conditions. The following models appear to work well in our experience and are close to the optimal design so that there is little room for improvement [34] [35]. A simple choice of working model for $\psi(x_j, a)$ is the power model

$$\psi(x_j, a) = \alpha_i^a, (i = 1, \dots, k) \tag{2.1}$$

where $0 < \alpha_1 < \dots < \alpha_k < 1$ and $0 < a < \infty$. The function introduced in the original CRM model by O’Quigley, Pepe and Fisher [1] is an increasing, convex-concave curve, the hyperbolic tangent function

$$\psi(x_j, a) = [(\tanh x_j + 1)/2]^a$$

with $a > 0$. One parameter logistic function was also proposed [25]:

$$\psi(x_j, a) = \frac{\exp(a_0 + ax_j)}{1 + \exp(a_0 + ax_j)}$$

with the intercept a_0 fixed.

Figure B.4 illustrates the basic idea of the 1-parameter CRM model. The true mechanism generating the observations could be much away from the working model overall. However, close to the target toxicity rate, the true situation and the working model coincide fairly well [1]. In order to decide the appropriate level at which to treat a patient, we need to estimate the probability of the toxic response at each level. These can be either Bayesian or likelihood based. The Bayesian estimator was used in the original paper by O’Quigley, Pepe, and Fisher [1]. Under the Bayesian framework, we can set up a prior for the parameter a and update the parameter distribution after observing data. However, we would currently recommend use of likelihood CRM (CRML) based on likelihood theory [34]. In order to be able to maximize the log-likelihood on the interior of the parameter space, we need heterogeneity among the response, i.e. at least one toxic and one non-toxic response [33]. In the context of a “pure likelihood” framework, we need a two-stage design with an initial escalation scheme to provide some heterogeneity in the response. This stage can use a standard Up-and-Down approach. We

can also view this first stage as collecting empirical data. With this data we can proceed to the second stage, the CRM modeling stage. Information from clinical investigation prior to the clinical trial and other considerations can be utilized in establishing the starting dose level and the initial escalation scheme. The two-stage design can come under the heading of an empirical Bayesian design. Therefore, both the Bayesian design and the two-stage design can be grouped under the broad Bayesian umbrella. In this thesis, if not otherwise specified, all CRM models are under the two-stage design.

After j patients complete treatment and heterogeneity already appears in the dataset, we can feed in the dataset, and get the maximum likelihood estimate \hat{a} of the parameter a . \hat{a} maximizes the following log-likelihood function

$$\log L(a) = \sum_{i=1}^j y_i \cdot \log \psi(x_i, a) + (1 - y_i) \cdot \log(1 - \psi(x_i, a))$$

With \hat{a} , $\hat{R}(d_i)$ ($i = 1, \dots, k$), the estimates of the true toxicity rate $R(d_i)$, ($i = 1, \dots, k$) at the k dose levels can be calculated. The recommended dose for the $(j + 1)$ th patient is the one satisfying

$$|\hat{R}(x_{j+1}) - \theta| < |\hat{R}(d_i) - \theta| \quad (2.2)$$

where $i = 1, \dots, k, x_{j+1} \neq d_i$ and θ is the target toxicity rate for the study. This recommended dose is the MTD for the trial if the j th patient is the last patient for the trial.

We use an example to demonstrate the implementation of the CRM model after heterogeneity appears. In this example, there are 6 pre-specified doses and the target rate is 0.20. Sixteen patients have completed the treatment and their

treatment outcomes are displayed in Table A.2. Under the one-parameter power model $\psi(d_i, a) = \alpha_i^a$, the log-likelihood can be expressed as:

$$\log L(a) = \log(1 - \alpha_1^a) + 4\log(1 - \alpha_2^a) + 8\log(1 - \alpha_3^a) + 2\log(\alpha_3^a) + \log(\alpha_4^a)$$

The maximum likelihood estimate of the parameter a , is 1.345. The 3rd dose with an estimated toxicity rate of 0.198 is the recommended dose for next patient or is the MTD if the study is completed (Table A.3).

2.2 One-parameter versus Two-parameter

As stated in Chapter 1, the goal of early phase dose-finding trials is to find the MTD and the CRM proposed by O'Quigley, Pepe, and Fisher [1] is a well-accepted model-based method. In the original CRM work, the authors studied both one- and two-parameter models with the one-parameter model as the main focus. Refer Section 2.1 for the one-parameter CRM for the standard one-group response model. The true mechanism generating the observations could deviate from the one-parameter working model overall, however, close to the target toxicity rate, the true situation and the working model coincide fairly well [1]. We need to bear in mind that the primary goal of phase I trials is to locate the MTD. Therefore, only the accuracy of the toxicity estimate of the dose closest to the target rate matters most. The statistical properties of the one-parameter CRM have been well studied [32] [33] [36].

The most used two-parameter CRM is the logistic model with intercept a_0 and slope a_1 . The probability of a toxicity from the administration of dose x is denoted

$R(x)$. Often the logarithm transformation of x is used. The relationship between $R(x)$ and x can be expressed as:

$$\log \frac{R(x)}{1 - R(x)} = a_0 + a_1 \log x \quad (2.3)$$

or

$$R(x) = \frac{\exp(a_0 + a_1 \log x)}{1 + \exp(a_0 + a_1 \log x)} \quad (2.4)$$

where the intercept a_0 and the slope a_1 are two unknown parameters. Some researchers claimed that a two-parameter CRM can capture the whole dose-toxicity curve [37] [38]. It is true that a two-parameter logistic function can produce a dose-response curve. However, due to the special feature of data from the CRM design, the predicted dose-response curve may not represent the true dose-response relationship. In the early stage of a phase I clinical trial in oncology, the sparsity of data does not enable accurate estimation of the two unknown parameters. By the end of a trial, due to the design and ethic restrictions of the trials, patients in the study are concentrated onto a single dose, the estimated MTD. The lowest doses and the highest doses could rarely or never be used, resulting in poor estimation of the two parameters. The estimated parameters would have wide standard errors, which leads to a non-informative dose-toxicity curve [39]. On the other hand, similarly to the one-parameter CRM, the predicted DLT probabilities are fairly accurate at the target dose.

Another concern for the two-parameter CRM is that priors for a_0 and a_1 are necessary to begin modeling. It is difficult to have a 2-stage design as with the one-parameter model. It is difficult for the first stage to collect heterogeneity suffi-

cient to estimate both parameters. The prior distributions for unknown parameters can be in a complex mathematic form. However, implementation of a study design using this type of prior is of difficulty. One possible way of expressing the priors is through pseudo-data, or imaginary data. Next two sections show us why we should choose one-parameter over two-parameter.

2.2.1 Lack of Stability of the Two-parameter CRM

Consider the following situation which corresponds to the set-up used in Gerke and Siedentop [40]. We put a pseudo-data prior consisting of a single patient at level one in which there is one third of a patient suffering a DLT. At the last level, say level 6, we put one half of a patient suffering a DLT and one half of a patient tolerating the treatment. This is the prior which is used in ADEPT, a statistical software implementing the logistic model CRM [41]. Suppose that the first entered patient suffers a DLT. For the usual one-parameter CRM, if one of the levels must be chosen and we are not allowing experimentation outside the range, then the recommended level for patient 2 would be level 1. Also, regardless of the level at which we treat the first entered patient, the recommended level for patient 2 is level 1. For the two-parameter CRM, the recommended level for patient 2 is level 6! The reason for this is not hard to understand and the design can be patched up. But it is not reassuring. This example illustrates the erratic or very noisy behavior of the two-parameter CRM for those patients entered early in the study, especially at small sample sizes. Furthermore, it takes longer than we might imagine for things to stabilize. In our simulations, we often see patient assignments similar to those in Table A.4. The true underlying toxicity rates are 0.35, 0.45, 0.55, 0.70, 0.80, 0.95 for doses 0.6, 1.2, 2.0, 3.0, 4.0, 5.0 mg. The

true MTD is the very first dose while the two-parameter CRM recommended the very last one (the target rate is 0.33). In another simulation for a scenario with 10 pre-specified doses, a target toxicity 0.33 and the same prior described above, the first patient had no toxicity. The two-parameter model recommended the *8th* dose for the second patient. The true MTD is the first dose and the *8th* dose would produce toxicity with a probability of 0.85. The two-parameter model will assign the second patient to a too high dose regardless of the outcome of the first patient. While it is true that the Bayesian prior does bring some stability, under the workable priors which are in use and in particular those used by ADEPT, there remains instability. The reason is not to be found within the inferential setting but is an immediate consequence of the two-parameter model. The two-parameter model is too flexible for the limited data and it even cannot guarantee the non-decreasing dose-toxicity relationship. The erratic behavior may cause more early patients to be assigned to high doses.

Consider Figure B.5 and Figure B.6 contrasting two approaches: the first one based on the two-stage one-parameter CRM, and the second based on the two-parameter CRM with a pseudo-data prior as described in the above paragraph. The target toxicity rate is 0.33 and Table A.5 indicates the true rates generating the data. These are chosen so that each of all six doses corresponds to the MTD in one of the scenarios. The cohort size is 1. On the basis of 2000 simulations, we indicate the percentage of the time that any given level is used for the first 8 and 4 patients. The poor behavior of the two-parameter CRM is immediately apparent. It assigns many more patients to higher dose levels due to the fact that the estimated second parameter (the slope) can be negative, zero or without bound so that the higher doses can appear much safer than they really are. In

all cases, except the last one it is outperformed by the two-stage one-parameter CRM. Looking at the first 4 patients produces yet more unfavorable results for the two-parameter CRM. An even larger proportion of patients are assigned to unsafe and untested doses compared to the first 8 patients. Figure B.7 shows that the 2-parameter CRM consistently yields more toxicities over the first 4 or 8 patients for the same target dose. The two-parameter CRM is simply a more aggressive design than the one-parameter CRM due to its lack of stability. Additionally, priors are necessary for the two-parameter CRM to get started. However, the consequence of mis-specified priors can be disastrous. Even if the priors are weak, it could be influential when the number of patients in a trial is small.

After the end of a trial, one may ignore the prior, estimate the parameters using the data from the trial (maximum likelihood estimates), and decide the final MTD recommendation. Very often, the data does not have enough information (“heterogeneity”) to estimate two parameters as in the logistic model. This happens especially when the number of testing doses is small. Another concern is the MTD recommendation based on ignoring the prior can be very different from the one recommended based on including the prior. Let MTD_{prior} be MTD based on including priors and MTD_{obs} be MTD based on observed data only. The nature of CRM design causes patients accumulated on a single dose, MTD_{prior} . There are few or no patients treated on MTD_{obs} when MTD_{prior} and MTD_{obs} are different. The toxicity estimation for dose MTD_{obs} is far from accurate due to the great imprecision of parameter estimation. Actually, the observed data can have good toxicity prediction only for dose MTD_{prior} .

2.2.2 Performance of the Two-parameter CRM with Pseudo-data for Shallow Dose-toxicity Curves

By their nature and amount, priors express both the position of the parameters (a_0, a_1) and the strength of that opinion. The effect of the priors is equivalent to extra “imaginary” data. One possible way of expressing prior opinion is through consideration of pseudo-data [26]. Essentially, some imaginary data are created as shown in Table A.6. In principle, the expert has a free choice in constructing this table. The values of t_{-1} and t_0 can be non-integer. n_{-1} and n_0 can be non-integer as well. However, with non-integer numbers of patients, it is more difficult to implement in software. The popular pseudo-data have n_{-1} “imaginary” patients on the lowest and n_0 on the highest doses. $p_{-1}n_{-1}$ patients on the lowest dose and p_0n_0 on the highest dose have toxicity. p_{-1} and p_0 are the prior toxicity rates for the lowest and the highest doses. The default pseudo-data used by Whitehead and Brunier are displayed in Table A.7. p_{-1} is often set as the target rate and p_0 as 0.50 [39] [40]. This pseudo-data express the belief that the lowest dose is the target dose and puts the 1st patient on this dose. The combined pseudo-data and real data are used to estimate the posterior distribution of (a_0, a_1) . The values of (a_0, a_1) that maximize the posterior distribution (modal estimates) are the same as maximum likelihood estimates of the combined pseudo and real data. This makes it feasible to fit the 2-parameter CRM model using frequentist logistic regression software such as SAS PROC LOGISTIC. Next cohort of patients will be assigned to the dose whose predicted toxicity rate from Equation (2.4) is closest to the target.

Whitehead [39] stated that even with $n_{-1} = 3$ and $n_0 = 3$, the prior opinion is weak and the real data will soon overcome and dominate the imaginary find-

ings from six subjects. This statement is optimistic for the situations where the treatments are very safe, that is, the true MTD is the highest dose. Comparing with the one-parameter CRM, the two-parameter CRM performs poorly for these situations as shown by simulations.

We randomly picked up a few typical safe scenarios when the highest dose among the pre-specified ones is the true MTD. Table A.8 shows these scenarios ranging from 3 doses to 7 doses. The target toxicity rate is 0.20 and a total of 24 patients are used. The pseudo-data for these simulations are shown in Table A.9. The dosages d_i for the two-parameter model are 0.20, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75 mg for $i = 1$ to 7. If a scenario has k doses, its dosages are the first k doses from above. For example, if $k = 3$, the dosages are 0.20, 0.25, 0.50. The logarithm transformation of dosage is used in the modeling. For the one-parameter CRM model with a two-stage design we used the power model with $\alpha_i, i = 1, \dots, 7$ as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7. The cohort size is one. On the basis of 2000 simulations, we indicated the percentage of times that any given level is recommended as MTD and the percentage that any given level is used for patients.

Figures B.8 and B.9 display the results contrasting the one-parameter CRM and the two-parameter CRM. In order to compare the performance more effectively, we plotted the cumulative percentage of the recommended MTD and the cumulative patient assignment starting from the highest dose, the true MTD. For all 5 scenarios, the lines for the one-parameter CRM always lie above those for the two-parameter one. This shows the two-stage one-parameter CRM outperforms the two-parameter in all cases. The superiority of one-parameter is even more striking when the number of doses in the test is small. For the 3-dose

scenario, 70% of the time, the one-parameter CRM chose the true MTD while the two-parameter CRM only did 32% of the time. As a direct result of more efficiently locating the right MTD, the one-parameter CRM put many more patients on or around the true MTD.

Only two “imaginary” patients are included for the logistic model and the priors are fairly weak. If 6 “imaginary” patients with 3 on d_{-1} or d_0 , it is anticipated that the performance of the logistic model would be further negatively affected.

In summary, we generally do not suggest using the two-parameter CRM model. In research published by Paoletti and Kramar [42], the authors concluded that a one-parameter power model outperforms a two-parameter logistic model, all the more as sample sizes increase.

Chapter 3

Two-group CRM Shift Model

In this chapter, we introduce the detail of the CRM shift model for the two-group situations including model specification and implementation in a likelihood framework. Other comparative CRM models to handle two groups are also covered.

3.1 Scheme I: CRM Shift model

3.1.1 A Motivating Example

To illustrate and provide motivation for pursuing the method described in this section, we consider a real clinical dose-finding study for breast cancer treatment. This is a multi-center phase I/II study of dasatinib plus capecitabine for paclitaxel-refractory metastatic breast cancer patients and dasatinib plus fulvestrant for hormone-sensitive, progressive metastatic breast cancer patients. The goal of the phase I of this study is to determine the MTD of dasatinib. We expect patients receiving capecitabine chemotherapy to be more sensitive to dasatinib

and thus to have a higher probability of toxicity compared with those receiving fulvestrant hormone therapy. The MTD of dastinib for the hormone therapy patients should be higher than or at least as high as the MTD for the chemotherapy patients. Additionally, the clinicians for this study believe that the MTD elevation for the hormone therapy group will not exceed 2 levels on the scale of the pre-specified doses. For example, if the MTD for the chemotherapy group is the 2nd dose, the MTD for the hormone therapy will not be dose 5 or higher. This is a motivating example for us to study the CRM shift model for heterogeneous populations.

3.1.2 CRM Shift Model Detail

In the CRM shift model, we keep the first parameter to model the dose-response relationship for one group and set up restriction for the difference between the two groups. The difference takes values from a small finite set. If the first group has identified dose level d_0 as the best recommended dose, then the other group will be recommended with either the same level or some level, one, two or more, steps away from d_0 . The direction and number of steps shifted will be set up based on the study setting and clinical background which can be provided by clinicians. The idea of the CRM shift model is to parameterize these steps directly.

In the CRM shift model, the indices for doses themselves are modeled.

$$R(d_i) = P(Y = 1|X = d_i) = E(Y|d_i) = \psi(d_{\phi(i)}, a); \quad \phi(i) = i + zh(i) \quad (3.1)$$

where $i = 1, \dots, k$ and

$$h(i) = \Delta I(1 \leq i + \Delta \leq k) + kI(i + \Delta > k) + I(i + \Delta < 1), \quad \Delta = -k + 1, \dots, k - 1. \quad (3.2)$$

z is an indicator variable for group, k is the number of the pre-specified doses, Δ is the levels shifted and I is an indicator function. Δ denotes the number of doses shifted for group 2 ($z = 1$) from group 1 ($z = 0$). A positive Δ indicates the toxicity rates of group 2 (G_2) shift upwards and the MTD shifts downwards, relative to group 1 (G_1). Similarly, a negative Δ means the toxicity rates of G_2 shift downwards and the MTD shifts upwards, relative to G_1 . The zero value of Δ shows the two groups have the same MTD. It is easy to add a discrete prior on Δ . If for example, the toxicity order between the two groups is of no doubt, we can restrict Δ to be only non-positive or only non-negative. We can also give the most weight to $\Delta = 0$ and only allow one or two dose levels shifted if the evidence of the accumulating data points strongly in that direction.

If we use a power model, we can model the toxicity profile as:

$$\psi(d_i, a) = \alpha_{\phi(i)}^a, \quad i = 1, \dots, k, \quad \phi(i) = i + zh(i) \quad (3.3)$$

where $0 < \alpha_m < \alpha_n < 1$ for $m < n$, $0 < a < \infty$ and z and $h(i)$ are defined as above in equation 3.1. If the order is known, the sign of $h(i)$ (or Δ) is known. When the $z = 0$ group is more sensitive to the treatment, $h(i)$ is less than or equal to 0. If we further know the possible number of doses shifted, we can restrict $h(i)$ (or Δ) to an even shorter list.

Strictly speaking, the CRM shift model has two parameters, Δ and a . The statistical modeling involves two steps: selecting Δ from a short list and producing

the posterior distribution of parameter a under a Bayesian framework or calculating the maximum likelihood estimate (MLE) of a under a likelihood framework. In this thesis, we only demonstrate the implementation of the shift model under a likelihood framework. As mentioned before, there is no critical difference between Bayesian framework and likelihood framework in implementing the CRM model. In Bayesian framework, we need to specify prior for the parameter a , while in likelihood framework, we need a two-stage design to collect empirical prior for the parameter a .

A two-stage design involves two stages. The first stage can be an up-and-down design until heterogeneity appears across both groups (both groups have patients when heterogeneity appears) and the second stage is a CRM shift model stage where data are used to estimate the parameter and to calculate the toxicity rates for the $2k$ doses. Denote the data for the first n_1 patients from G_1 as $\{(x(1)_1, y(1)_1), \dots, (x(n_1)_1, y(n_1)_1)\}$ and the first n_2 patients from G_2 as $\{(x(1)_2, y(1)_2), \dots, (x(n_2)_2, y(n_2)_2)\}$. The likelihood function for n_1 patients from G_1 and n_2 from G_2 is

$$L(a_1, a_2) = \prod_{j=1}^{n_1} \psi_1(x(j)_1, a_1)^{y(j)_1} [1 - \psi_1(x(j)_1, a_1)]^{(1-y(j)_1)} \prod_{j=1}^{n_2} \psi_2(x(j)_2, a_2)^{y(j)_2} [1 - \psi_2(x(j)_2, a_2)]^{(1-y(j)_2)}$$

Under the CRM shift model, the toxicity rate at d_i for G_1 is assumed to be the same as the toxicity rate at $d_{i+\Delta}$ for G_2 . Therefore, we can recode the data from G_2 as $\{(x(1)_{\frac{\Delta}{2}}, y(1)_2), \dots, (x(n_2)_{\frac{\Delta}{2}}, y(n_2)_2)\}$, where the dose index ($i = 1, 2, \dots, k$) of $x(j)_2$ is the index of $x(j)_{\frac{\Delta}{2}}$ minus Δ . A single function ψ with the parameter a can

be used for both groups and the likelihood can be rewritten as

$$L_{\Delta}(a) = \prod_{j=1}^{n_1} \psi(x(j)_1, a)^{y(j)_1} [1 - \psi(x(j)_1, a)]^{(1-y(j)_1)} \\ \prod_{j=1}^{n_2} \psi(x(j)_2^{\Delta}, a)^{y(j)_2} [1 - \psi(x(j)_2^{\Delta}, a)]^{(1-y(j)_2)}$$

and the log-likelihood as

$$\log L_{\Delta}(a) = \sum_{j=1}^{n_1} y(j)_1 \log \psi(x(j)_1, a) + [1 - y(j)_1] \log [1 - \psi(x(j)_1, a)] \\ + \sum_{j=1}^{n_2} y(j)_2 \log \psi(x(j)_2^{\Delta}, a) + [1 - y(j)_2] \log [1 - \psi(x(j)_2^{\Delta}, a)]$$

Below is an example to demonstrate implementation of the CRM shift model. There are 6 candidate doses for the two groups. Clinicians believe G_2 patients can tolerate the treatment better or at least not worse than G_1 patients and MTD for G_2 can shift upwards up to 2 levels. After n_1 patients from G_1 and n_2 from G_2 have completed the treatment, we can write the log-likelihood with a power function as:

$$\log L_{\Delta}(a) = \sum_{i=1}^6 n_{1i(1)} a \log(\alpha_i) + n_{1i(0)} \log(1 - \alpha_i^a) + \sum_{i=1}^6 n_{2i(1)} a \log(\alpha_{i+\Delta}) + n_{2i(0)} \log(1 - \alpha_{i+\Delta}^a) \quad (3.4)$$

where Δ (with potential values of 0, -1, -2) is the levels shifted, $n_{1i(1)}$ and $n_{1i(0)}$ are numbers of DLT and non-DLT from G_1 on dose d_i , and $n_{2i(1)}$ and $n_{2i(0)}$ are numbers of DLT and non-DLT from G_2 on dose d_i . Also

$$\sum_{i=1}^6 (n_{1i(1)} + n_{1i(0)}) = n_1, \quad \sum_{i=1}^6 (n_{2i(1)} + n_{2i(0)}) = n_2, \text{ and } n_1 + n_2 = n$$

We write out the detail to demonstrate the design idea.

For $\Delta = 0$

$$\begin{aligned}\log L_0(a) &= \sum_{i=1}^6 n_{1i(1)} a \log(\alpha_i) + n_{1i(0)} \log(1 - \alpha_i^a) \\ &+ \sum_{i=1}^6 n_{2i(1)} a \log(\alpha_i) + n_{2i(0)} \log(1 - \alpha_i^a) \\ &= \sum_{i=1}^6 (n_{1i(1)} + n_{2i(1)}) a \log(\alpha_i) + (n_{1i(0)} + n_{2i(0)}) \log(1 - \alpha_i^a)\end{aligned}$$

For $\Delta = -1$

$$\begin{aligned}\log L_{-1}(a) &= \sum_{i=1}^6 n_{1i(1)} a \log(\alpha_i) + n_{1i(0)} \log(1 - \alpha_i^a) \\ &+ \sum_{i=1}^6 n_{2i(1)} a \log(\alpha_{i-1}) + n_{2i(0)} \log(1 - \alpha_{i-1}^a)\end{aligned}$$

For $\Delta = -2$

$$\begin{aligned}\log L_{-2}(a) &= \sum_{i=1}^6 n_{1i(1)} a \log(\alpha_i) + n_{1i(0)} \log(1 - \alpha_i^a) \\ &+ \sum_{i=1}^6 n_{2i(1)} a \log(\alpha_{i-2}) + n_{2i(0)} \log(1 - \alpha_{i-2}^a)\end{aligned}$$

The maximum log-likelihoods for the above three situations corresponding to Δ value of 0, -1 or -2 are calculated and the one with the highest maximum log-likelihood is chosen to estimate the parameter a . If we denote by $(\hat{\Delta}, \hat{a})$ values of (Δ, a) maximizing the equation (3.4) after the inclusion of n patients, then the

estimated toxicity rates at dose d_i are $\alpha_i^{\hat{\alpha}}$ and $\alpha_{i+\hat{\Delta}}^{\hat{\alpha}}$ for G_1 and G_2 , respectively. If the $(n + 1)$ th patient belongs to G_1 , this patient will receive a dose level that minimizes $|\alpha_i^{\hat{\alpha}} - \theta|$ with $i \in \{1, \dots, k\}$. On the other hand, if the $(n + 1)$ th patient belongs to G_2 , the recommended dose level minimizes $|\alpha_{i+\hat{\Delta}}^{\hat{\alpha}} - \theta|$. If n th patient is the last one for the study, the above recommended doses are the two MTDs for the two groups.

The real escalation is case-specific and is subject to ethical requirement [31]. It has two operationally conflicting guidelines. The first is that we need to proceed cautiously and to not put an unacceptably large number of patients at risk for toxic side effects. The second one is we would like to reach the target dose with as few patients as possible, thereby avoiding treating too many patients at doses too far below the MTD. In the first stage, the two groups can escalate independently, which usually is not efficient. We can also combine the two groups if the order is not obvious. Or we can enforce the order information. One example of adding the order information is as shown in a typical trial history displayed in Figure B.10. The cohort size is 1 for both stages and there are 6 pre-specified doses in this example. Suppose patients from G_2 can tolerate treatment better than those from G_1 . With a design where dose allocation in G_2 must be at least equal to the dose currently allocated to G_1 , then G_1 cannot skip any dose while G_2 can if some patients (from any group) have been treated on that dose level. For example, the 2nd patient is the 1st patient from G_2 and is assigned to the dose level 2 instead of 1 since one patient from G_1 has been treated on the dose level 1 and there is no toxicity. However, the 3rd patient is from G_1 and assigned to the dose level 2 even though 1 patient from G_2 has been treated without toxicity. As soon as there is heterogeneity present, with toxicity and non-toxicity across both groups, we can

move to the second stage and patient assignments are based on the probability estimates from the CRM shift model.

3.2 Scheme II: Two-parameter CRM Model

In O'Quigley, Shen and Gamst [30], the authors introduced a second parameter to the one-group CRM to model two-group dose-finding studies. The new parameter measures the direction and magnitude of the difference between the two groups. The dose-toxicity curves are modeled as

$$P(Y = 1|x, I = 1) = \psi_1(x, a),$$

$$P(Y = 1|x, I = 2) = \psi_2(x, a, b),$$

where (a, b) is a pair of one-dimensional parameters. This specification shows the fact that the two groups share information in terms of reaction to the treatment. The parameter a indicates the shared information and b distinguishes the two groups. If we use the power model as shown in the equation (2.1), the probability of toxicity of dose d_i can be specified as

$$P(Y = 1|X = d_i : z) = \psi(d_i, a, bz) = \alpha_i^{a+bz}, (i = 1, \dots, k) \quad (3.5)$$

where $0 < \alpha_m < \alpha_n < 1$ for $m < n$, $a > 0$ and $a + bz > 0$. z is an indicator for the group with 0 for G_1 and 1 for G_2 . As usual in the CRM context, the models are mis-specified and not assumed to provide an accurate global fit to the true

dose-toxicity curves. O'Quigley and Paoletti [31] extended the above model for studies in which we anticipate some ordering between groups. The same model is used with the restriction of sign of parameter b to indicate the direction of group ordering. $b \geq 0$ shows we anticipate G_2 is at least as tolerant as G_1 to the treatment and its MTD should be no lower than the MTD of G_1 . Similarly, $b \leq 0$ shows we anticipate G_2 is at least as sensitive as G_1 to the treatment and its MTD should be no higher than MTD of G_1 . The specific $b = 0$ corresponds to the absence of patient heterogeneity. In order to estimate the two parameters, one can use a Bayesian estimate or maximum likelihood estimate based on the first n_1 patients from G_1 and n_2 patients from G_2 ($n_1 + n_2 = n$). The log-likelihood based on the first n patients can be written as:

$$\begin{aligned} \log L(a, b) &= \sum_{i=1}^k n_{1i(1)} a \log(\alpha_i) + n_{1i(0)} \log(1 - \alpha_i^a) \\ &+ \sum_{i=1}^k n_{2i(1)} (a + b) \log(\alpha_i) + n_{2i(0)} \log(1 - \alpha_i^{a+b}) \end{aligned}$$

where $n_{1i(1)}$ and $n_{1i(0)}$ are the numbers of toxicities and non-toxicities from G_1 patients on dose i , respectively; $n_{2i(1)}$ and $n_{2i(0)}$ are the numbers of toxicities and non-toxicities from G_2 patients on dose i , respectively. If we denote the maximum likelihood estimates of (a, b) as (\hat{a}, \hat{b}) , then the estimated dose-toxicity relations are $\psi(x, \hat{a})$ and $\psi(x, \hat{a} + \hat{b})$ for the two groups. The next patient from either group will be allocated a dose whose estimated toxicity is closest to the target.

In the likelihood framework, the 2-stage design is applied. After heterogeneity appears in both groups, we can start to use the two-group two-parameter CRM model to estimate the toxicity rate $R(d_i)$ for both groups. If the order is obvious between populations, which is often true, we can use the two-group two-parameter

CRM model for ordered groups.

3.3 Scheme III: Two Groups, Two Separate Trials

With this scheme, we carry out two independent one-sample studies and apply the one-sample CRM for both studies separately. The dose allocation for one group does not rely on information from the other group. The implementation of the procedure is the same as explained in O'Quigley and Shen [34].

3.4 Scheme IV: Ignore the Group Difference and Pool the Two Groups

In this scheme, the group difference is ignored and the two groups are pooled together. The one-sample CRM is then applied to the combined group. Clearly, this method should provide the best performance when the two groups have the same or similar dose-toxicity curves around the target dose. When the two groups have different target doses (therefore different dose-toxicity curves), we expect the recommended dose would be some average between the two target doses. Again the detailed implementation of the one-sample CRM can be found in O'Quigley and Shen [34].

Chapter 4

Two-group CRM Model Performance

In this chapter, we study the performance of the four CRM designs for two ordered groups by simulations. We also show the enhanced performance by prior information for the levels shifted and the nearly optimal performance by comparing with the optimal design. For all simulations, we use a 2-stage design and pure likelihood-based inference based on power functions.

4.1 Comparison of Four Schemes for Two-group Studies

In this section, we study the performance of the four CRM designs for two ordered groups by simulations. For all simulations and designs, we use a 2-stage design and pure likelihood-based inference based on power functions. For the CRM shift model, we assume $\Delta \in (0, -1, -2)$. We specify a few different true dose response scenarios to generate the responses of simulated patients (Table

A.10). These scenarios represent the true Δ equals to 0, -1 , -2 , or -3 . The number of doses (k) is 6. Based on clinical background, patients from G_2 can tolerate treatment at least as well as those from G_1 and the maximum dose shift is 2 ($\Delta \in (0, -1, -2)$). Table A.11 shows the working codes for the power models. Code 0 (0.20, 0.30, 0.50, 0.70, 0.80, 0.90) is used for the two-parameter CRM model and the one-sample CRM including schemes II, III and IV. For the shift model, the working codes are code 0 (0.20, 0.30, 0.50, 0.70, 0.80, 0.90) for G_1 and for G_2 when $\Delta = 0$, code 1 (0.10, 0.20, 0.30, 0.50, 0.70, 0.80) for G_2 when $\Delta = -1$ and code 2 (0.05, 0.10, 0.20, 0.30, 0.50, 0.70) for G_2 when $\Delta = -2$. Thirty-two patients are randomly assigned to the two groups with 16 for each group or 24 for G_1 and 8 for G_2 . The target toxicity rate is $\theta = 0.20$. Each trial is replicated 5000 times. All schemes have an escalation stage and a CRM stage. In the escalation stage, the group size is 1 and the first patient receives the first dose. In the two-group CRM shift model, if a dose has one patient from G_1 treated and there is no toxicity, the next patient from either group is assigned to the next higher dose. If a dose has one patient from G_2 treated and there is no toxicity, the next patient from G_1 cannot skip this dose. After heterogeneity appears (among all treated patients) across both groups, the CRM shift model comes into play. For the two-group two-parameter design, before heterogeneity appears in any group, the escalation strategy is the same as for the shift model. When we have heterogeneity among the responses in one group, for example G_1 , we can proceed in G_1 using the usual 1-sample CRM design. For G_2 , escalation continues according to the ordering requirement. Once we encounter heterogeneity in G_2 , we have complete heterogeneity and we can fit a two-parameter model to the recorded data.

Tables A.12, A.13, A.14, A.15 show the results of simulations including dis-

tribution of the recommended MTD, and distribution of assigned doses. For instance Table A.12 records 0.27 in the cell from column of d_1 and row of “Prop(MTD), Scheme I”. This value indicates 27% of the times the CRM shift design recommends dose 1 as the MTD for G_1 . The value of 0.35 in the cell from column of d_1 and row of “Prop(Pat), Scheme I” indicates on average the CRM shift model assigns 35% of the 16 patients from G_1 on dose 1. The bold columns are for the proportion that the true MTD is recommended by the study or the proportion of patients who are assigned on the true MTD.

In scenario A, the two study groups are identical. With no surprise, the combined 1-sample CRM design (scheme IV) performs best. The performance of the CRM shift model (scheme I) is nearly the same as the combined 1-sample CRM design. The 2-group 2-separate study CRM design (scheme III) performs worst. The effective sample size for scheme III is only half of that for scheme IV. The 2-group 2-parameter CRM design (scheme II) performs slightly worse than the CRM shift model, however, the difference is small. Only 4% of the time, the CRM shift model recommends an MTD 2 or more levels away from the true MTD for G_1 and 15% of the time for G_2 . On average, the chance that the CRM shift model recommends MTD as the true MTD or one dose away is the same as that of the one-combined sample CRM design. When the accrual ratio is 0.75/0.25 for G_1/G_2 , the order of performance remains the same as the equal accrual situations.

As expected, the combined 1-sample CRM performs poorly in the presence of heterogeneity. The design recommends an MTD which is too toxic for G_1 while it has suboptimal efficacy for G_2 . When the true number of levels shifted is within the speculated range for the shift model (scenarios B and C), the CRM shift model

performs best. This is also true if the accrual is not equal. As a direct consequence, the CRM shift model allocates more patients onto the most promising doses. If the true number of levels shifted is out of the speculated range (scenario D), the 2-group 2-parameter CRM model outperforms the CRM shift model.

4.2 Using Prior Information for the Levels Shifted

We can add prior probability for the number of doses shifted. For example, the clinicians may not only know that the MTD for the G_2 patients shifts upward no more than 2 levels ($\Delta = 0, -1, -2$), but also may know there is a greater chance that $\Delta = -1$. In this case, if we utilize this piece of information, we will have a higher chance of recommending the right MTD for both groups. We use scenario B as an example for detail (Table A.16). With higher prior confidence in the correct number of doses shifted, the CRM shift model has better performance. When the prior confidence for $\Delta = -1$ increases to 50%, the chance to find the right MTD increases by 7.5% (prior = 0.17, 0.50, 0.33). That is a substantial increase. If we have 60% prior confidence for $\Delta = -1$, we gain 8.5% more chance to recommend the right MTD.

4.3 Comparison of the CRM Shift Model and the Optimal Design

The purpose of phase I clinical trials and the implemented phase I designs is to identify a level whose associated probability of toxicity is as close to a specified

target as possible. Researchers develop models in hope to have improved performance. To answer the question whether there is some limit beyond which further improvements can not be achieved without making strong parametric assumptions about the shape of the dose-toxicity curve, O'Quigley *et al.* [35] proposed the concept of optimal design. The design is not available for practical implementation and exists only as the theoretical benchmark. Taking into account the limited sample size of the phase I oncology trials and the particular toxicity probabilities at various levels, the optimal design can tell us how well we might be able to do and the performance limit a design might be able to reach.

In an optimal design, we have all toxicity information for any given patient at all doses [35]. In other words, the highest tolerated dose level for each patient is known to us. This is not practically possible but it can be constructed in the context of a theoretical or virtual investigation. In a real study, each subject provides only partial or incomplete information. Therefore, the performance of any practical design is limited to the performance of the optimal design for the same treatment.

We can define an optimal design using the concept of complete and incomplete information.

4.3.1 Incomplete Information

An optimal design utilizes complete information which is typically unavailable in reality. In a real study, each subject provides only partial or incomplete information. The assumption of monotonicity of dose-toxicity response relationship implies that if a subject experiences a toxicity at dose d_i , this subject will have a

toxicity at higher doses. However, there is not enough information about this subject's toxicity response at lower doses. And if a subject experiences no toxicity at dose d_i , it tells that this subject will have no toxicity at lower doses, but it doesn't provide enough information about whether this patient will have toxicity at higher doses. As an illustration, we assume there are 6 doses and suppose that subject h experiences a toxicity at dose d_4 and subject j has no toxicity at dose d_2 . This is summarized in Table A.17 where a * indicates incomplete information. Any given patient on treatment dose k will provide incomplete information as shown in Table A.17.

4.3.2 Complete Information

All toxicity information for any given patient at all doses is available in an optimal design. To understand this virtual design, we can imagine there are m (m = the number of doses) identical clones for each subject with each clone assigned on each dose. Then we can present the complete data as in Table A.18 below. Compared to the incomplete information table, we have complete information about each subject at each dose. In other words, the highest tolerated dose level is known to us. The empirical estimates of the binomial proportions are fully efficient for this optimal design.

In order to generate the complete information in simulations, a random generator will be used to generate a uniform random variables T_j . For each realized value t_j of T_j , it is compared with the pre-specified true toxicity rate R_i at dose i . If $t_j < R_i$, subject j has toxicity at dose i or $T_{ji} = 1$; otherwise, subject j has no toxicity at dose i or $T_{ji} = 0$. With n subjects in study, the toxicity rate for each

dose from the optimal design can be calculated as:

$$\hat{R}_i = \frac{1}{n} \sum_{j=1}^n T_{ji} \quad (4.1)$$

The dose with \hat{R}_i closest to the target toxicity rate is the recommended MTD. With repeated simulation, we can calculate the proportion of a dose being recommended as the MTD.

4.3.3 Near Optimality of the CRM Shift Model Performance

We use scenarios A and B with equal prior probabilities for $\Delta=0, -1$ or -2 to demonstrate the performance of the CRM shift model comparing to the optimal design. We carried out 5000 simulations of the two designs. Tables A.19 and A.20 show the recommended MTD distribution when the toxicity target rate is 0.20. $p_i(16)$ is the proportion of times that the shift model recommend dose level i as the MTD and $q_i(16)$ the analogous quantity for the optimal design. Figures B.11 and B.12 display the cumulative frequency of distance from the target for scenarios A and B. The distribution of the recommended MTD by the CRM shift model is nearly the same as that from the optimal designs. The CRM shift model not only outperforms the other CRM designs, but has nearly optimal performance. This suggests that very little can be gained if extra information is incorporated into dose-finding methods for heterogeneous populations.

Chapter 5

Three-group CRM Shift Model

In the previous 2 chapters, we focus on the 2-group CRM shift model. The ideas of the CRM shift model can be easily applied to 3 groups. Extension to more than 3 groups is theoretically plausible, however in practice, the implementation could be complicated. This chapter covers the situations where the 3-group CRM shift model can be applied to and the ideas and implementation of the 3-group CRM shift model.

In real practice, a prognostic factor could have 3 levels. Also patients could be divided into 3 groups by more than 1 prognostic factor. For example, besides the concomitant therapies that could affect sensitivity to the study drug, the level of prior treatment might cause different response to the new treatment. It is wise to take into account most important prognostic factors that will affect patients' sensitivity to the treatment under study. In the example for demonstration, we will consider two prognostic factors, concomitant treatment (chemotherapy versus hormone therapy) and the level of prior treatment (light versus heavy). There are four possible combinations of these two binary factors: light+hormone (LH),

light+chemo (LC), heavy+hormone (HH), and heavy+chemo (HC). Assume the dose-toxicity profile is the same for light+chemo and heavy+hormone. Therefore, there are 3 groups indeed.

We will demonstrate the implementation of the 3-group CRM shift model and compare its performance with the design of 3 separate studies. First we need to set up the model and model assumptions.

As in the 2-group CRM shift model, the indices for doses themselves are modeled in the 3-group CRM shift model.

$$R(d_i) = P(Y = 1|X = d_i) = E(Y|d_i) = \psi(d_{\phi(i)}, a); \quad \phi(i) = i - z_1 h_1(i) + z_2 h_2(i) \quad (5.1)$$

where $i = 1, \dots, k$ and

$$h_1(i) = \Delta_1 I(1 \leq i + \Delta_1 \leq k) + k I(i + \Delta_1 > k) + I(i + \Delta_1 < 1), \quad \Delta_1 = -k+1, \dots, k-1. \quad (5.2)$$

and

$$h_2(i) = \Delta_2 I(1 \leq i + \Delta_2 \leq k) + k I(i + \Delta_2 > k) + I(i + \Delta_2 < 1), \quad \Delta_2 = -k+1, \dots, k-1. \quad (5.3)$$

z_1 and z_2 are indicator variables for groups G_1 and G_3 , and k is the number of pre-specified doses for each group. $-\Delta_1$ is the level of toxicity shifted for G_1 ($z_1 = 1$) from G_2 ($z_1 = 0$). In other words, Δ_1 is the level of toxicity shifted for G_2 from G_1 . Δ_2 is the level of toxicity shifted for G_3 ($z_2 = 1$) from G_2 ($z_2 = 0$). I is an indicator function. A positive Δ_1 (or Δ_2) indicates the toxicity rates of G_2 (or G_3) shift upwards and the MTD shifts downwards, relative to G_1 (or G_2). Similarly, a negative Δ_1 (or Δ_2) means the toxicity rates of G_2 (or G_3) shift downwards and the

MTD shifts upwards, relative to G_1 (or G_2). The zero value of Δ_1 (or Δ_2) shows G_2 (or G_3) and G_1 (or G_2) have the same MTD. It is easy to add a discrete prior on Δ_1 and/or Δ_2 . If for example, the toxicity order between two groups is of no doubt, we can restrict Δ_1 and Δ_2 to be only non-positive or only non-negative. We can also give the most weight to $\Delta_1 = 0$ or $\Delta_2 = 0$ and only allow one or two dose levels shifted if the evidence of the accumulating data points strongly in that direction.

If we use a power model, we can model the toxicity profile as:

$$\psi(d_i, a) = \alpha_{\phi(i)}^a, \quad i = 1, \dots, k, \quad \phi(i) = i - z_1 h_1(i) + z_2 h_2(i) \quad (5.4)$$

where $0 < \alpha_m < \alpha_n < 1$ for $m < n$, $0 < a < \infty$ and $z_1, z_2, h_1(i)$ and $h_2(i)$ are defined as above in equation 5.1. If the order is known, the sign of $h_1(i)$ (or Δ_1) or $h_2(i)$ (or Δ_2) is known.

For the example of the prognostic factors of concomitant treatment and level of prior treatment, we denote:

1. Δ'_1 = Dose index of the MTD of Light+Chemo or Heavy+Hormone (G_2) - dose index of the MTD of Heavy+Chemo (G_1)
2. Δ'_2 = Dose index of the MTD of Light+Hormone (G_3) - dose index of the MTD of Light+Chemo or Heavy+Hormone (G_2)
3. Δ'_3 = Dose index of the MTD of Light+Hormone (G_3) - dose index of the MTD of Heavy+Chemo (G_1)

The above Δ'_1 and Δ'_2 are the opposites of Δ_1 and Δ_2 defined in (5.1).

Assume that one prognostic factor can cause the MTD to shift downwards up

to 1 level, we have $\Delta'_1 \in (0, 1)$, and $\Delta'_2 \in (0, 1)$. Table A.21 shows the levels shifted between groups. The value of Δ'_3 depends on Δ'_1 and Δ'_2 and assume the effects of the two factors are additive. Then we have $\Delta'_3 \in (0, 1, 2)$. There are 4 possible situations:

1. S1: if $\Delta'_1=0$ and $\Delta'_2=0$, then $\Delta'_3 = 0$
2. S2: if $\Delta'_1=0$ and $\Delta'_2=1$, then $\Delta'_3 = 1$
3. S3: if $\Delta'_1=1$ and $\Delta'_2=0$, then $\Delta'_3 = 1$
4. S4: if $\Delta'_1=1$ and $\Delta'_2=1$, then $\Delta'_3 = 2$

In a two-stage design, the 3-group CRM shift model starts to play after heterogeneity appears across the 3 groups (all 3 groups have patients when heterogeneity appears). As in the two-group CRM shift model, data are fitted into the competing likelihood functions corresponding to different situations determined by Δ'_1 , Δ'_2 and Δ'_3 . The likelihood function with the highest likelihood is used to estimate the single parameter and to calculate the toxicity rates for the $3k$ doses. Denote the data for the first n_1 patients from G_1 as $\{(x(1)_1, y(1)_1), \dots, (x(n_1)_1, y(n_1)_1)\}$, the first n_2 patients from G_2 as $\{(x(1)_2, y(1)_2), \dots, (x(n_2)_2, y(n_2)_2)\}$ and the first n_3 patients from G_3 as $\{(x(1)_3, y(1)_3), \dots, (x(n_3)_3, y(n_3)_3)\}$. The likelihood function for n_1 patients from G_1 , n_2 patients from G_2 and n_3 patients from G_3 is

$$L(a_1, a_2, a_3) = \prod_{j=1}^{n_1} \psi_1(x(j)_1, a_1)^{y(j)_1} [1 - \psi_1(x(j)_1, a_1)]^{(1-y(j)_1)} \\ \prod_{j=1}^{n_2} \psi_2(x(j)_2, a_2)^{y(j)_2} [1 - \psi_2(x(j)_2, a_2)]^{(1-y(j)_2)} \\ \prod_{j=1}^{n_3} \psi_3(x(j)_3, a_3)^{y(j)_3} [1 - \psi_3(x(j)_3, a_3)]^{(1-y(j)_3)}$$

Under the 3-group CRM shift model, the toxicity rate at d_i for G_1 is the same as the toxicity rate at $d_{i-\Delta'_1}$ for G_2 and is the same as $d_{i-\Delta'_3}$ for G_3 . Therefore, we can recode the data from G_2 as $\{(x(1)_2^{\Delta'_1}, y(1)_2), \dots, (x(n_2)_2^{\Delta'_1}, y(n_2)_2)\}$, where the dose index ($i = 1, 2, \dots, k$) of $x(j)_2$ equals to the index of $x(j)_2^{\Delta'_1}$ plus Δ'_1 . The data from G_3 can also be recoded as $\{(x(1)_3^{\Delta'_3}, y(1)_3), \dots, (x(n_3)_3^{\Delta'_3}, y(n_3)_3)\}$, where the dose index ($i = 1, 2, \dots, k$) of $x(j)_3$ equals to the index of $x(j)_3^{\Delta'_3}$ plus Δ'_3 . And the likelihood function can be rewritten as

$$L_{\Delta_1, \Delta_3}(a) = \prod_{j=1}^{n_1} \psi(x(j)_1, a)^{y(j)_1} [1 - \psi(x(j)_1, a)]^{(1-y(j)_1)} \prod_{j=1}^{n_2} \psi(x(j)_2^{\Delta'_1}, a)^{y(j)_2} [1 - \psi(x(j)_2^{\Delta'_1}, a)]^{(1-y(j)_2)} \prod_{j=1}^{n_3} \psi(x(j)_3^{\Delta'_3}, a)^{y(j)_3} [1 - \psi(x(j)_3^{\Delta'_3}, a)]^{(1-y(j)_3)}$$

To assess the performance of the 3-group CRM shift model, we use the scenario in Table A.22 for simulation. The performance is compared with the design of 3 separate studies. We demonstrate two situations: equal probability of shift and non-shift and unequal probability of shift and non-shift. The alpha values for the working model are the same as used for the 2-group CRM shift model simulations A.11 and the target rate is 0.20. When the enrollment is equal across the 3 groups, there are 16 patients for each group. When the enrollment is unequal, G_2 has 24 patients and G_1 and G_3 each have 12 patients. All simulations use a 2-stage design.

5.1 Equal Probability of Shift and Non-shift

If both Δ'_1 and Δ'_2 have the same prior probabilities for their two possible values, that is, $P(\Delta'_1 = 0) = P(\Delta'_1 = 1) = 0.5$ and $P(\Delta'_2 = 0) = P(\Delta'_2 = 1) = 0.5$, then the prior probabilities for S1-S4 are all equal to 0.25.

Simulation results are shown in Tables A.23 and A.24 to compare the shift model design with the 3 separate study design ignoring the order information between the 3 study groups.

The shift model enhances the performance greatly compared with the 3 separate study design, especially for G_1 and G_2 . The shift model picked up the correct MTDs with a much higher chance for G_1 and G_2 than the design of 3 separate studies (63% vs. 49% for G_1 and 63% vs. 51% for G_2). The percentage of correct MTD for G_3 is similar between the shift model design and the 3 separate study design (49% vs. 48%). However, there is a much higher chance the 3rd dose is chosen by the shift model design (42%) than by the 3 separate study design (27%). On the contrary, the 3 separate study design has a much higher chance to select a too toxic dose as the MTD. Nineteen percent of the time the 3 separate study design chooses the fifth dose (vs. only 4% by the shift model). The toxicity rate for the 5th dose is 0.39, which is way too high for a MTD. As a direct consequence of a higher chance to choose the correct MTDs, the shift model assigns more patients in G_1 and G_2 to their MTDs and more patients in G_3 to less toxic doses comparing with the design of 3 separate groups.

The advantage of the shift model design is even more obvious when the enrollment is differential (Table A.24), with more patients in G_2 . The performance of the shift model design is not affected by the unequal enrollment across all 3

groups (63% vs. 61%, 63% vs. 65% and 49% vs. 48% for G_1 , G_2 , and G_3 , respectively). More patients enrolled in G_2 enables the 3 separate study design to increase the performance of G_2 (from 51% to 65%). However, the performance of G_1 and G_3 is greatly dampened (49% vs. 43% and 48% vs 40% for G_1 and G_2 , respectively).

5.2 Unequal Probability of Shift and Non-shift

Assume the prior probabilities for $\Delta'_1 = 1$ and $\Delta'_2 = 1$ are both 0.80. In other words, it is believed that there is a higher chance each prognostic factor shifts the MTD upwards by 1 level than there is no effect on MTD. Then the prior probability for the 4 situations are: 0.04, 0.16, 0.16 and 0.64 for S1, S2, S3 and S4, respectively.

The simulation results are displayed in Tables A.23 and A.24. The more prior information on the true number of levels shifted increases the shift model performance dramatically across all 3 groups. Overall, the prior information enhances the performance by around 16% across the 3 groups. The largest increase is for G_3 and is nearly 40% (from 49% to 68%). The difference in performance becomes even larger between the shift model and the design with 3 separate studies (69% vs. 49%, 70% vs. 51% and 68% vs. 48% for the 3 groups with equal enrollment). Clearly, the utilization of the common information between groups and more confidence on the levels shifted not only greatly increase the chance to find the correct MTD but also provide a better chance to treat patients with potentially beneficial doses.

Chapter 6

Stopping Rules for the CRM Shift Model

The CRM is a sequential design and a fixed sample size n is decided upon for the study. It may be felt that phase I trials should be halted early whenever possible to quickly proceed to the phase II trials. This is especially true in oncology since the patient accrual is slow and phase II studies of response are lengthy. In an attempt to deal with these concerns, researchers have proposed some stopping rules to allow the early termination of phase I studies [1] [22] [43]. One of the peculiarities of the CRM is the property of convergence. It is shown in Shen and O'Quigley [33] that if the chosen working model satisfies certain conditions, then the method will converge almost surely to the level where the toxicity is the closest to the targeted rate. This thesis also has proved that under certain conditions, the two-group CRM shift model converges (Chapter 8). If given enough time, any trial will present a similar picture: early oscillation followed by a more or less long plateau. The goal is to stop the study at some earlier point of the

plateau. The rule suggested in the original CRM article [1] decides on the basis of a confidence interval whether enough precision has been reached and the study can be terminated. This rule was further studied by Heyd and Carlin [43]. O'Quigley and Reiner [22] proposed a new stopping rule based on the idea that continuing the study would not lead to a dose recommendation change with high probability. However, this rule, based on precision probabilistic calculation, is not straightforward to implement. A much simpler rule was suggested by O'Quigley [44] based on the idea of having settled at some level. In this rule, a study stops when the dose recommended to the next patient has already been allocated m times, where m is some number fixed at the beginning of the trial. This rule has the advantage of great simplicity and it also works fairly well in practice. We will use $m = 6$ for examples in this thesis.

In the context of heterogeneity, additional problems arise. The question of which rule (those used for the one-group cases) to use is not in the scope of this thesis. We will focus on the unique problem of two-group studies: stop the two groups separately or simultaneously.

In the first stopping rule (SR1), the two groups can stop at different times. If one group has 6 consecutive patients on one dose and the same dose is recommended to the next patient ($6 + 1$ rule), this group meets the stopping rule. No more patients will be recruited from this group. The other group will continue recruiting patients until it meets the same $6 + 1$ rule. Data from both groups are used to estimate the model parameter even after one group has halted recruiting.

In the second stopping rule (SR2), the two groups stop at the same times. Both groups must meet the $6 + 1$ rule simultaneously. When the study stops, one group has at least 6 consecutive patients on its MTD and the other group has

6 consecutive patients on its MTD. Clearly, more patients are required for SR2 than SR1. With more patients enrolled, studies with SR2 would have better performance overall than with SR1. However, the question is whether the gained performance is worthy of the extra patients enrolled and the extended study period. We perform simulations using a typical scenario to study the concern.

Simulations are performed to compare SR1 and SR2 for Scenario B in Table A.10. The simulation settings are similar to those for the performance simulation of the two-group CRM shift models.

- Maximum sample size: 36 per group.
- Target rate: 0.20.
- Number of simulations: 5000
- Working codes: refer to Table A.11
- Prior probabilities for levels shifted (Δ): 0.33 : 0.33 : 0.33 for $\Delta = 0, -1, -2$

The simulation shows that the median sample sizes using SR1 are 15 and 16 for groups 1 and 2 respectively versus 20 and 21 using SR2 (Table A.25). SR2 stops the study about 32% later than SR1. The associated accuracy gain in finding the right MTDs is only 5% (Table A.26). In order to proceed to the phase II of the experimental drug, SR1 is preferred in this simulated study. For studies in practice, the study team should perform simulations to decide which stopping rule is preferred and specify it prior to the trial.

The better performance of SR1 in the above example could be caused by the fact that one MTD deviates from the other one by a certain level and this deviation is captured in the shift model. Once one MTD is determined, the relationship

between the the MTDs leads to the determination of the other MTD. Extra patients cannot improve the accuracy of the recommended MTDs by a significant amount.

Chapter 7

Robustness of the CRM Shift Model

One natural question a reader would ask is the robustness and practical utility of the two-group CRM model. Since, in truth, a classification into two groups may be a rough summary of many groups or even a continuum, it is important to study (most likely via the tool of simulation) how this affects estimation of the MTD. Another important question is that of misclassification. If a certain percentage of subjects are incorrectly assigned to a particular group status then, clearly, if this percentage is very high, there is little hope of accurate identification of the MTD. For smaller percentages though things may still work well enough and the question is how much error is tolerable.

For all simulations in this section, 16 patients per group are used and 5000 simulations are performed. The target toxicity rate is 0.20. The working codes for the power model are listed in Table A.11.

7.1 The Effect of Classification

For many situations where the prognostic factors differentiating the population are categorical, there is no need to concern about the effect of classification. Typical examples are hormone therapy resistance (resistant vs. sensitive) and gender (male vs. female). For many characteristics of a population, it is hard or even impossible to classify the populations into two groups. For example, it is a difficult task to correctly classify patients into adults versus children since there are many different age thresholds for the definition of adulthood and patient biological development varies. Another example is to classify the degree of prior treatment. We will use this as our example in our simulation work.

In order to carry out simulations, we set up representative scenarios. Let us assume the effect of degree of prior treatment (a continuum) on patient toxicity towards the new treatment follows a logistic function:

$$R(i, x) = \frac{\exp(a_i + \beta x)}{1 + \exp(a_i + \beta x)}$$

where $\beta \geq 0$, x is degree of prior treatment, a_i measures the toxicity of dose d_i on patients with no prior treatment and β measures the effect of prior treatment on sensitivity to the new treatment. The variable x is the degree of prior treatment and ranges between 0 (minimal or no prior treatment) and 1 (maximal prior treatment). The question of how to quantify the prior treatment is not in the scope of this thesis. The distribution of x is assumed to be uniform to simplify illustration. If the differential degree of prior treatment is ignored, that is, the population is

treated as homogenous, then the population average toxicity rate is:

$$R(i) = \frac{1}{\beta} \log \frac{\exp(a_i + \beta) + 1}{\exp(a_i) + 1}$$

for $\beta > 0$. When $\beta = 0$,

$$R(i, x) = \frac{\exp(a_i)}{1 + \exp(a_i)}$$

If we classify the degree of prior treatment into light ($x \leq 0.50$) and heavy ($x > 0.50$) with a threshold of 0.50. The average toxicities for the two groups (when $\beta > 0$) are:

$$R(i)_l = \frac{2}{\beta} \log \frac{\exp(a_i + 0.5\beta) + 1}{\exp(a_i) + 1}$$

and

$$R(i)_h = \frac{2}{\beta} \log \frac{\exp(a_i + \beta) + 1}{\exp(a_i + 0.5\beta) + 1}$$

If we classify the degree of prior treatment into light as $x \leq 0.66$ and heavy as $x > 0.66$. Then the two group averages (when $\beta \neq 0$) are:

$$R(i)_l = \frac{3}{2\beta} \log \frac{\exp(a_i + 0.66\beta) + 1}{\exp(a_i) + 1}$$

and

$$R(i)_h = \frac{3}{\beta} \log \frac{\exp(a_i + \beta) + 1}{\exp(a_i + 0.66\beta) + 1}$$

We will examine the effect of different classifications of the prior treatment on the performance of the two-group CRM shift design by studying the above two classifications for two scenarios, with scenario 1: weak effect of the degree of prior treatment and scenario 2: strong effect of the degree of prior treatment. We choose $a_i = -5, -3, -1.8, -1, -0.2, 0.5$, and $\beta = 0.50$ for scenario 1 and $\beta = 2.0$ for scenario 2. The average toxicity rates are shown in Table A.27.

There is no surprise that classification has little effect on model performance when the effect of prior treatment is weak (Tables A.28 and A.29). For scenario 1, the toxicity rates for patients with no or minimal prior treatment ($x = 0$) are 0.01, 0.05, 0.14, 0.27, 0.45, and 0.62 and the rates for those with maximal prior treatment ($x = 1$) are 0.01, 0.08, 0.21, 0.38, 0.57, and 0.73 for doses 1 to 6. The population average toxicity rates are 0.01, 0.06, 0.18, 0.32, 0.51, and 0.68. The toxicity rates do not differ significantly between the two classifications ($x = 0.50$ or $x = 0.66$). For weak scenarios, classification does not change the MTDs since the MTDs for minimal prior treatment and maximal prior treatment do not differ much.

The effect of classification is stronger when the effect of a prognostic factor on patient's sensitivity to treatment is strong. Classifying the prior treatment at the threshold of 0.66 instead of 0.50 does increase the toxicity rate of the true MTDs even though the MTD locations remain the same (Tables A.30 and A.31). The toxicity rates for patients with no or minimal prior treatment ($x = 0$) are 0.01, 0.05, 0.14, 0.27, 0.45, and 0.62 and the rates for those with maximal prior treatment ($x = 1$) are 0.05, 0.27, 0.55, 0.73, 0.86, and 0.92 for doses 1 to 6. The population average toxicity rates are 0.02, 0.13, 0.32, 0.50, 0.68, and 0.80. The MTD for $x = 0$ is the 3rd or the 4th dose and the MTD for $x = 1$ is the 2nd. The overall MTD for population is the 2nd dose. Classification with no extreme threshold

(x value away from 0 or 1) produces the light prior treatment group with the true MTD of the 3rd dose and the heavy prior treatment group with the true MTD of the 2nd dose. Therefore, the number of levels shifted between two MTDs is 1 for the 2nd scenario with a reasonable classification. For situations where the prognostic effect is so strong that the number of levels shifted can be large, different classification will produce different MTD locations. Regarding which classification is optimal, it depends on how much the study team concern about population heterogeneity and whether any conventional threshold exists. Overall, classifying a heterogeneous population into two less heterogeneous groups leads to more accurate MTD estimation as shown in Chapter 4.

7.2 The Effect of Misclassification

Misclassification causes each group to be less homogenous within the group. The group difference is weakened, which leads to attenuation of the group effect on the MTD. Clearly, the performance of the two-group CRM shift model with misclassified patients will be between the one-group CRM model ignoring group difference and the two-group CRM shift model with accurate classification. We used scenarios B and C in Table A.10 as illustration examples. The simulation settings were the same as for Section 4.1:

- Number of patients per group: 16.
- Target rate: 0.20.
- Number of simulations: 5000
- Working codes: as shown in Table A.11

- Prior probabilities for levels shifted (Δ): 0.33 : 0.33 : 0.33 for $\Delta = 0, -1, -2$

With increased misclassification, the MTD recommended by the study moves up for the sensitive group and moves down for the less sensitive group (Tables A.32 and A.33). When the true number of levels shifted is one as in scenario B, the effect of misclassification is less than when the true number of levels shifted is two as in scenario C. The performance of scenario B does not change much when misclassification rate rises from 2% to 50%. For scenario C, the performance is highly dampened by the increased misclassification. With 20% or more misclassification, the 3rd dose is recommended with the highest chance as the MTDs for both groups. With 50% misclassification, the chance to recommend the 3rd dose reaches nearly 50% for both groups.

In conclusion, when the group difference is large, the performance of the 2-group CRM shift model is strongly attenuated by misclassification. Therefore, accurately classifying patients to a right group is crucial in boosting the study performance, especially when the prognostic factors are strong.

Chapter 8

Consistency of the CRM Shift Model

This chapter of the thesis is to examine the asymptotic convergence of the 2-group CRM shift model under some practically realistic conditions. The goal is to show convergence to the correct MTDs in probability, in both groups, when sample sizes for both groups are allowed to increase without bound. In practice, we are inevitably faced with small-to-moderate sample sizes so that large sample properties alone will not be wholly convincing. However, the result of asymptotic consistency will be helpful in that it can not only provide some assurance as to the basic statistical soundness of the model but can also provide guidance in the choice and structure of different candidate working models. Convergence arguments will be obtained from considerations of the likelihood. The same arguments apply to Bayesian estimation as long as the prior is other than degenerate.

Shen and O'Quigley have proved the asymptotic consistency of the classic one-sample CRM model [33]. The two-sample two-parameter CRM model has also been proved consistent under some minor assumptions [30]. Our simulations have shown the recommended MTDs from the two-group CRM shift design

converge quickly to the true MTDs.

In this chapter we shall study the asymptotic behavior of the estimate of the parameter and the recommended dose levels and show that the sequentially designed experiments based on the two-group CRM shift model consistently select the true model and the true MTDs.

8.1 The Set-up

This section explains the notation of functions, parameters etc. and the expression of log likelihood for the 2-sample CRM shift model.

Let X_1 and X_2 , taking values d_1, \dots, d_m , stand for the dose levels for the two groups used for an experiment, and let Y_1 and Y_2 be the binary response for the experiment. $Y_1 = 1$ or $Y_2 = 1$ stands for DLT observed and $Y_1 = 0$ or $Y_2 = 0$ for no DLT observed. Assume that the true dose-toxicity function generating the data is given by

$P(Y_1 = 1/X_1 = x_1) = R_1(x_1)$ and $P(Y_2 = 1/X_2 = x_2) = R_2(x_2)$. Denote x_{10} and x_{20} as the true MTDs for G_1 and G_2 . One is interested in estimating the dose levels x_{10} for G_1 and x_{20} for G_2 at which $R_1(x_{10})$ and $R_2(x_{20})$ equal to a target probability, θ . It should be noted, however, that $R_1(x_{10})$ and $R_2(x_{20})$ do not have to equal to θ exactly.

Let $\psi_1(x_1, a_1)$ and $\psi_2(x_2, a_2)$ be the two one-parameter functions to model the dose-toxicity relations for the two groups. $\psi_1(\cdot, a_1)$ and $\psi_2(\cdot, a_2)$ are two increasing functions of x and a_1 and a_2 are unknown parameters. Here the models $\psi_1(x_1, \cdot)$ and $\psi_2(x_2, \cdot)$ are rich enough so that for each x_1 and $p \in (0, 1)$, there exists a

value a_1 such that

$$\psi_1(x_1, a_1) = p \quad (8.1)$$

and for each x_2 and $p \in (0, 1)$, there exists a value a_2 such that

$$\psi_2(x_2, a_2) = p \quad (8.2)$$

In other words, as long as there exists a dose level that gives the target probability of toxicity we can always estimate this target level by first estimating the curve determined by (8.1) or (8.2).

Denote the data for the first n_1 patients from G_1 as $\{(x(1)_1, y(1)_1), \dots, (x(n_1)_1, y(n_1)_1)\}$ and the first n_2 patients from G_2 as $\{(x(1)_2, y(1)_2), \dots, (x(n_2)_2, y(n_2)_2)\}$.

The likelihood function for n_1 patients from G_1 and n_2 from G_2 together is

$$L(a_1, a_2) = \prod_{j=1}^{n_1} \psi_1\{x(j)_1, a_1\}^{y(j)_1} [1 - \psi_1\{x(j)_1, a_1\}]^{1-y(j)_1} \prod_{j=1}^{n_2} \psi_2\{x(j)_2, a_2\}^{y(j)_2} [1 - \psi_2\{x(j)_2, a_2\}]^{1-y(j)_2}$$

Denote D as the set containing potential values of Δ , the number of toxicity levels shifted for G_2 from G_1 (or the opposite of the number of MTD levels shifted for G_2 from G_1). The set D includes a finite number of integer values. Let Δ_0 be the value of Δ satisfying $x_{10} = x_{20} + \Delta_0$. Δ_0 is the true number of toxicity levels shifted for G_2 from G_1 .

Under the two-group CRM shift model, values of $\psi_1(d_i, a_1)$ are shifted from

values of $\psi_2(d_i, a_2)$ and they share one common unknown parameter, a . That is,

$$\psi_1(d_i, a_1) = \psi(d_i, a)$$

and

$$\psi_2(d_i, a_2) = \psi(d_{i+\Delta}, a)$$

where Δ is the number of toxicity levels shifted. For simplicity, denote $\psi(d_{i+\Delta}, a)$ as $\psi_\Delta(d_i, a)$ for individual Δ . Strictly speaking, there are two unknown parameters for the 2-sample CRM shift model: Δ , the number of toxicity levels shifted and a , the unknown parameter for function ψ . With a particular Δ , the log-likelihood function can be rewritten using a single dose-toxicity function ψ with the parameter a :

$$\begin{aligned} \log L_\Delta(a) &= \sum_{j=1}^{n_1} y(j)_1 \log \psi(x(j)_1, a) + [1 - y(j)_1] \log[1 - \psi(x(j)_1, a)] \quad (8.3) \\ &+ \sum_{j=1}^{n_2} y(j)_2 \log \psi_\Delta(x(j)_2, a) + [1 - y(j)_2] \log[1 - \psi_\Delta(x(j)_2, a)] \end{aligned}$$

The joint log-likelihood function of a and Δ can be expressed as a function of a for each Δ due to the fact that Δ is discrete and finite. Each possible Δ has a corresponding log-likelihood function and a related estimated $\hat{a}_n(\Delta)$ (where $n = n_1 + n_2$). The Δ and corresponding \hat{a}_n from the log-likelihood formula that produces the highest maximum log-likelihood will be used to calculate the best current estimates of the dose-toxicity relationships and therefore the next recommended dose levels for the two groups.

The maximum likelihood procedure begins once we have observed some het-

erogeneity in the response across both groups. Define

$$n_0 = \inf\{m : 0 < \sum_{i=1}^{m_1} y(i)_1 + \sum_{j=1}^{m-m_1} y(j)_2 < m, m_1 > 0, m - m_1 > 0\}$$

n_0 is the first time that both outcomes (DLT and non-DLT) of the response are observed and both groups have patients and therefore the maximum likelihood estimate can be calculated.

Denote a_0 as the parameter a value for (8.3) satisfying $\psi(x_{10}, a) = \theta$. If $R_1(x_{10}) = R_2(x_{20})$, then

$$\psi(x_{10}, a_0) = R_1(x_{10})$$

and

$$\psi(x_{20}^\Delta, a_0) = R_2(x_{20})$$

where the dose index of x_{20}^Δ equals to the index of x_{20} plus Δ . If $R_1(x_{10}) \neq R_2(x_{20})$, but both are close to the target θ , then

$$\min(R_1(x_{10}), R_2(x_{20})) < \psi(x_{10}, a_0) = \psi(x_{20}^\Delta, a_0) < \max(R_1(x_{10}), R_2(x_{20}))$$

That is, the estimated toxicity rates for the two MTDs based on the model will be between the two true rates.

8.2 Specifying Conditions

Choosing a particular model $\psi(x, a)$ is largely a matter of personal taste. It is most convenient for the sake of asymptotics to focus on $\psi(x, a)$ satisfying the following conditions.

Condition M1. (a) For each a , functions $\psi(\cdot, a)$ is strictly increasing.

(b) Function $\psi(x, \cdot)$ is continuous and is strictly monotone in a in the same direction for all x from the same group.

Condition M2. For each $0 < t < 1$ and each x , the functions

$$s(t, x, a) = t \frac{\psi'}{\psi}(x, a) + (1 - t) \frac{-\psi'}{1 - \psi}(x, a)$$

is continuous and is strictly monotone in a .

Condition M3. The parameter a belongs to a finite interval $[A, B]$. *Condition M3* is appropriate for the purpose of condition for (8.1) or (8.2) if we believe that the dose-toxicity functions $R_1(x)$ and $R_2(x)$ is bounded away from 0 and 1. There are many functions satisfying Conditions M1 and M2. One example is the power function used for simulations in this thesis.

Suppose that the actual dose-toxicity relations $R_1(x)$ and $R_2(x)$ satisfy the following.

Condition D1. The target dose level are x_{10} and x_{20} , that is $R_1(x_{10}) = R_2(x_{20}) = \theta$ where θ is the target rate.

Condition D2. The probabilities of toxicity at d_1, \dots, d_m satisfy $0 < R(d_1) < \dots < R(d_m) < 1$ for both groups separately.

This *Condition D2* translates the idea that the function $R(\cdot)$ is invertible on the interval $(0, 1)$.

Ideally, we would like that $\psi(d_i, a_0) \equiv R_1(d_i)$ and $\psi(d_i, a_0) \equiv R_2(d_i)$ for $i = 1, \dots, m$. It is unlikely for this to be true in real situations. In order to establish the asymptotic consistency for the recommended dose levels, we have to require that the actual dose-toxicity curves be not too far away from the assumed ones. To characterize their differences, introduce sets S_1 , S_2 and S .

$$S_1 = \{a : |\psi(x_{10}, a) - \theta| < |\psi(x_{1i}, a) - \theta|, \text{ for all } d_i \neq x_{10}\} \quad (8.4)$$

and

$$S_2 = \{a : |\psi_{\Delta_0}(x_{20}, a) - \theta| < |\psi_{\Delta_0}(x_{2i}, a) - \theta|, \text{ for all } d_i \neq x_{20}\} \quad (8.5)$$

Where $i = 1, \dots, m$.

The sets S_1 and S_2 consist of parameters of “good” curves for group 1 and group 2, respectively, according to which x_{10} and x_{20} are the two target dose levels for the two groups.

Define the common set S :

$$S = S_1 \cap S_2 \quad (8.6)$$

The set S consists of parameters of “good” curves for both groups according to which x_{10} and x_{20} are the target dose levels and $x_{10} = x_{20} + \Delta_0$. We would like to have S as large as possible. For $i = 1, \dots, m$, let a_{1i} be such that $\psi(d_i, a_{1i}) = R_{1i}$ and a_{2i} be such that $\psi(d_i, a_{2i}) = R_{2i}$. We need the following condition.

Condition D3. For $i = 1, \dots, m$, $a_{1i} \in S$, and $a_{2i} \in S$. This condition ensures that the assumed dose-toxicity relation is not drastically different from the actual

one and prevents the recommended next doses would be “stuck” in a dose other than x_{10} or x_{20} .

Condition D4. x_{20} shifts away from x_{10} by $-\Delta_0$ levels, and $\Delta_0 \in D$. This condition ensures that the potential values of Δ include the true value. In other words, the prior probability for Δ_0 is larger than 0.

A final condition: *Condition P1.* Let $p_n = n_1 n^{-1}$. For some $0 < p < 1$, $p_n \rightarrow p$ almost surely as $n \rightarrow \infty$. This assumption ensures that both groups are represented in the same sample and hence that no group completely dominates the sample.

M1-M3 and D1-D3 are the same conditions for the asymptotic consistency of the classic one-sample CRM [33]. Cheung and Chappell [45] indicate that, in general, there exists some working models that are more sensitive than others when closeness of the converged MTD estimate to the target dose is considered. Cheung and Chappell introduced the *indifference interval* of a dose finding design to measure such “closeness”. Indifference interval is defined as an interval in which the toxicity probability of the selected dose will eventually fall. The smaller the width of the indifference interval of a design, the higher resolution the design will have asymptotically. Cheung and Chappell also suggested that if the true dose-toxicity function is steep around the target level, for large enough n , the dose recommended by the CRM is the true MTD under a less restrictive condition that the indifference interval of the true MTD does not overlap with the indifference intervals of other doses.

8.3 Theorem

Assume that Conditions M1-M3, Conditions D1-D4 and Condition P1 are satisfied. For $n = n_1 + n_2$ large enough, let $\hat{\Delta}_n$ be the estimate of the true Δ_0 for the 2-group CRM shift model, and \hat{a}_n be the maximum likelihood estimate of the parameter a for the 2-group CRM shift model. And let $x_1(n+1)$ be the recommended dose level for the next experiment from G_1 and $x_2(n+1)$ for G_2 . Then as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, $\hat{\Delta}_n \rightarrow \Delta_0$, $\hat{a}_n \rightarrow a_0$, $x_1(n_1+1) \rightarrow x_{10}$ and $x_2(n_2+1) \rightarrow x_{20}$ in probability.

8.4 Proof of Theorem

The proof has two steps. First, we need to show the dose-toxicity function $\psi_{\Delta}(d_i, a)$ with $\Delta = \Delta_0$ is consistently selected over all possible values of $\Delta \in D$. Then we need to prove that under the correct model with Δ_0 which is essentially a one-group CRM model, the true MTDs, x_{10} and x_{20} , for the two groups are consistently selected. The second piece has been proved by Shen and O'Quigley [33]. Details will be provided in this thesis as well.

8.4.1 Consistent Model Selection

In this section, we shall show the model with Δ_0 consistently produces the highest maximum likelihood and therefore is selected consistently.

Denote:

1. ψ be the dose-toxicity function for the 2-group CRM shift model (as defined

before)

2. ψ_T be the true dose-toxicity function satisfying $\psi_T(d_i, a) = \psi(d_{i+\Delta_0}, a)$ for G_2
3. ψ_F be the competing dose-toxicity function satisfying $\psi_F(d_i, a) = \psi(d_{i+\Delta}, a)$ for G_2 when $\Delta \in D$ and $\Delta \neq \Delta_0$
4. $\log L_T$ be the log-likelihood for G_1 and G_2 together when $\Delta = \Delta_0$
5. $\log L_F$ be the log-likelihood for G_1 and G_2 together when $\Delta \neq \Delta_0$
6. $\log L_1$ be the log-likelihood for G_1 only
7. $\log L_{2T}$ be the log-likelihood for G_2 only, using function $\psi(d_{i+\Delta_0}, a)$
8. $\log L_{2F}$ be the log-likelihood for G_2 only, using function $\psi(d_{i+\Delta}, a)$, when $\Delta \neq \Delta_0$

Assume the toxicity rates at the two true MTDs are the same for the two groups (*Condition D1*). It should be noted that it is not necessary for $R_1(x_{10}) = R_2(x_{20})$ in order for the two-group CRM shift model to converge. This condition makes the proof easy and straightforward for readers to understand.

Then we have:

$$\begin{aligned} \log L_T = & \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, a) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, a)) \\ & + \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_T(x(j)_2, a) + (1 - y(j)_2) \cdot \log(1 - \psi_T(x(j)_2, a)) \end{aligned}$$

$$\begin{aligned} \log L_F &= \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, b) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, b)) \\ &\quad + \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b)) \end{aligned}$$

$$\log L_1 = \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, a_1) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, a_1))$$

$$\log L_{2T} = \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_T(x(j)_2, a_2) + (1 - y(j)_2) \cdot \log(1 - \psi_T(x(j)_2, a_2))$$

$$\log L_{2F} = \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b_2) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b_2))$$

Let \hat{a} , \hat{b} , \hat{a}_1 , \hat{a}_2 , and \hat{b}_2 be the maximum likelihood estimates of a , b , a_1 , a_2 , and b_2 in the above 5 log likelihood formulas. Based on the Theorem 1 from Shen and O'Quigley [33] (the convergence of the one-sample CRM model), with $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, \hat{a}_1 , \hat{a}_2 , and \hat{b}_2 converge to the true values in probability. That is, with $n_1 \rightarrow \infty$, \hat{a}_1 tends to a_1 and with $n_2 \rightarrow \infty$, \hat{a}_2 tends to a_2 and \hat{b}_2 tends to b_2 in probability.

Based on the two-sample CRM shift model, under the true Δ_0 , we have $a_1 = a_2 = a$. On the other hand, with n_1 and n_2 go to infinity, \hat{b} converges to b in probability, a value between a and b_2 . b is an unknown number with value between a and b_2 and the value is affected by the asymptotic ratio of n_1/n_2 . With the increase of n_1/n_2 , b gets closer to a and with the decrease of n_1/n_2 , b gets closer

to b_2 . Clearly, with n_1 and n_2 going to infinity, we have

$$\liminf(\log L_1(a) - \log L_1(b)) > 0 \quad (8.7)$$

and

$$\liminf(\log L_{2F}(b_2) - \log L_{2F}(b)) > 0 \quad (8.8)$$

since a (and $a = a_1$) and b_2 are the true parameters for functions ψ and ψ_F .

The following steps show $P(\log L_T(\hat{a}) > \log L_F(\hat{b})) \rightarrow 1$ as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$.

$$\begin{aligned} & \log L_T(\hat{a}) - \log L_F(\hat{b}) \\ &= \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, \hat{a}) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, \hat{a})) \\ &+ \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_{2T}(x(j)_2, \hat{a}) + (1 - y(j)_2) \cdot \log(1 - \psi_{2T}(x(j)_2, \hat{a})) \\ &- \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, \hat{b}) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, \hat{b})) \\ &- \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_{2F}(x(j)_2, \hat{b}) + (1 - y(j)_2) \cdot \log(1 - \psi_{2F}(x(j)_2, \hat{b})) \end{aligned}$$

Since as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, $\hat{a} \rightarrow a$ and $\hat{b} \rightarrow b$ in probability, then $\lim(P(\log L_T(\hat{a}) = \log L_T(a))) = 1$ and $\lim(P(\log L_F(\hat{b}) = \log L_F(b))) = 1$. Therefore,

$$\lim(P(\log L_T(\hat{a}) - \log L_F(\hat{b}) = \log L_T(a) - \log L_F(b))) = 1 \quad (8.9)$$

In the following steps, we want to show $\liminf (\log L_T(a) - \log L_F(b)) > 0$.

$\log L_T(a) - \log L_F(b)$ can be partitioned into 3 parts as shown below.

$$\begin{aligned}
& \log L_T(a) - \log L_F(b) \\
= & \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, a) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, a)) \\
& + \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_T(x(j)_2, a) + (1 - y(j)_2) \cdot \log(1 - \psi_T(x(j)_2, a)) \\
& - \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, b) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, b)) \\
& - \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b)) \\
= & \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, a) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, a)) \\
& - \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, b) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, b)) \\
& + \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_T(x(j)_2, a) + (1 - y(j)_2) \cdot \log(1 - \psi_T(x(j)_2, a)) \\
& - \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b_2) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b_2)) \\
& + \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b_2) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b_2)) \\
& - \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b))
\end{aligned}$$

The first part is

$$\begin{aligned}
 P1 &= \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, a) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, a)) \\
 &\quad - \sum_{j=1}^{n_1} y(j)_1 \cdot \log \psi(x(j)_1, b) + (1 - y(j)_1) \cdot \log(1 - \psi(x(j)_1, b)) \\
 &= \log L_1(a) - \log L_1(b)
 \end{aligned}$$

and the third part is

$$\begin{aligned}
 P3 &= \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b_2) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b_2)) \\
 &\quad - \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b)) \\
 &= \log L_{2F}(b_2) - \log L_{2F}(b)
 \end{aligned}$$

It follows from 8.7 and 8.8 that as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$,

$$\liminf P1 > 0$$

and

$$\liminf P3 > 0$$

The second part is

$$P2 = \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_T(x(j)_2, a) + (1 - y(j)_2) \cdot \log(1 - \psi_T(x(j)_2, a)) \\ - \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi_F(x(j)_2, b_2) + (1 - y(j)_2) \cdot \log(1 - \psi_F(x(j)_2, b_2))$$

and we need to show that as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$,

$$\liminf P2 \geq 0$$

Based on the notations, $\psi_T(d_i, a) = \psi(d_{i+\Delta_0}, a)$ and $\psi_F(d_i, b_2) = \psi(d_{i+\Delta}, b_2)$ where $\Delta \neq \Delta_0$. Therefore,

$$P2 = \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi(x(j)_2^{\Delta_0}, a) + (1 - y(j)_2) \cdot \log(1 - \psi(x(j)_2^{\Delta_0}, a)) \\ - \sum_{j=1}^{n_2} y(j)_2 \cdot \log \psi(x(j)_2^{\Delta}, b_2) + (1 - y(j)_2) \cdot \log(1 - \psi(x(j)_2^{\Delta}, b_2))$$

As mentioned in Section 8.1 (The Set-up), the dose-toxicity function ψ is rich enough so that for each d_i and $p \in (0, 1)$, there exists a value a such that $\psi(d_i, a) = p$. Therefore, for each d_i and $p \in (0, 1)$, there exist a value a and a value b_2 such that $\psi(d_{i+\Delta_0}, a) = p$ and $\psi(d_{i+\Delta}, b_2) = p$. In other words, $\psi(d_{i+\Delta_0}, a)$ and $\psi(d_{i+\Delta}, b_2)$ are equivalent in modeling the data from G_2 , and as $n_2 \rightarrow \infty$, $\lim P2 = 0$ (or $\liminf P2 = 0$).

The above steps show that

$$\liminf (\log L_T(a) - \log L_F(b)) = \liminf (P1 + P2 + P3) > 0$$

Together with the result of 8.9, we have

$$\lim P(\log L_T(\hat{a}) - \log L_F(\hat{b}) > 0) = 1.$$

as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$.

This has proved that with sufficiently large n_1 and n_2 , the maximum likelihood for the model with the true number of levels shifted is consistently the highest. In other words, as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$

$$\hat{\Delta}_n \rightarrow \Delta_0$$

in probability. And $\log L_T$ would be chosen to estimate the only parameter for the situation where $R_1(x_{10}) = R_2(x_{20})$. When $R_1(x_{10}) \neq R_2(x_{20})$ but with no dramatic difference, a_1 , a_2 and a are close. The maximum likelihood for $\hat{\Delta}_n = \Delta_0$ is the highest and this likelihood function would be chosen to estimate the parameter. When $R_1(x_{10})$ and $R_2(x_{20})$ differ substantially, the consistency of model selection might be in question. The performance of the two-group CRM shift model might be questionable. However, in this situation, the toxicity rates of the true MTDs for the two groups deviate from the target rate for a large amount such that the dose re-planning should be considered.

The next step shows under the correct model (in this case, it reduces to one-sample CRM model), the parameter estimate \hat{a}_n converges to a_0 , $x_1(n+1)$ to x_{10}

and $x_2(n+1)$ to x_{20} in probability.

8.4.2 Consistent Dose Selection

Under $\Delta = \Delta_0$, the two-group question is reduced to a one-group question. The “new” one-group study has $m+|\Delta_0|$ doses. Denote these levels as $d_1, \dots, d_m, d_{m+1}, \dots, d_{m+|\Delta_0|}$ if $\Delta_0 \geq 0$ or $d_{\Delta_0}, d_{\Delta_0+1}, \dots, d_1, \dots, d_m$ if $\Delta_0 < 0$. Write the dose-toxicity function as $P(Y = 1/X = x) = \psi(x, a)$. Table A.34 helps understand this notation.

Based on the Theorem 1 from Shen and O’Quigley [33], under certain conditions, the maximum likelihood estimate of the parameter for the one-sample CRM model converges to the solution to the equation $\psi(x_0, a) = R(x_0)$ where x_0 is the target dose; and the next recommended dose converges to x_0 . Our proof of the asymptotic properties of the “new” one-sample CRM model follows the same idea of the proof in Shen and O’Quigley [33].

Suppose in the “new” one -group study, the first n experiments are carried out at $x(1), \dots, x(n)$ with responses y_1, \dots, y_n . Define

$$I_n(a) = \frac{1}{n} \sum_{i=1}^n [y_i \frac{\psi'}{\psi}(x(i), a) + (1 - y_i) \frac{-\psi'}{1 - \psi}(x(i), a)], \quad (8.10)$$

$$\tilde{I}_n(a) = \frac{1}{n} \sum_{i=1}^n [R\{x(i)\} \frac{\psi'}{\psi}(x(i), a) + (1 - R\{x(i)\}) \frac{-\psi'}{1 - \psi}(x(i), a)] \quad (8.11)$$

Based on the *Condition* M1, for each dose level x_j , $y_i \frac{\psi'}{\psi}(x(i) = x_j, a)$ and $(1 - y_i) \frac{-\psi'}{1 - \psi}(x(i) = x_j, a)$ are uniformly continuous in a over the finite interval $[A, B]$.

Then for each x_j and any $\epsilon > 0$, there must exist a partition $A = t_0 < t_1 < \dots <$

$t_s = B$ such that for any $a \in [A, B]$, there is a $t_{s_0} \in [A, B]$ such that

$$\left| \frac{\psi'}{\psi}(x_j, a) - \frac{\psi'}{\psi}(x_j, t_{s_0}) \right| < \epsilon \quad (8.12)$$

$$\left| \frac{\psi'}{1-\psi}(x_j, a) - \frac{\psi'}{1-\psi}(x_j, t_{s_0}) \right| < \epsilon \quad (8.13)$$

Since there are only $m + |\Delta_0|$ dose levels, the partition may be chosen so that (8.12) and (8.13) are valid for all doses.

$I_n(a) - \tilde{I}_n(a)$ can be partitioned into 3 parts as shown below.

$$\begin{aligned} I_n(a) - \tilde{I}_n(a) &= I_n(a) - \frac{1}{n} \sum_{i=1}^n [y_i \frac{\psi'}{\psi}(x(i), t_{s_0}) + (1 - y_i) \frac{-\psi'}{1-\psi}(x(i), t_{s_0})] \\ &+ \frac{1}{n} \sum_{i=1}^n [y_i \frac{\psi'}{\psi}(x(i), t_{s_0}) + (1 - y_i) \frac{-\psi'}{1-\psi}(x(i), t_{s_0})] \\ &- \frac{1}{n} \sum_{i=1}^n [R\{x(i)\} \frac{\psi'}{\psi}(x(i), t_{s_0}) + (1 - R\{x(i)\}) \frac{-\psi'}{1-\psi}(x(i), t_{s_0})] \\ &+ \frac{1}{n} \sum_{i=1}^n [R\{x(i)\} \frac{\psi'}{\psi}(x(i), t_{s_0}) + (1 - R\{x(i)\}) \frac{-\psi'}{1-\psi}(x(i), t_{s_0})] \\ &- \frac{1}{n} \sum_{i=1}^n [R\{x(i)\} \frac{\psi'}{\psi}(x(i), a) + (1 - R\{x(i)\}) \frac{-\psi'}{1-\psi}(x(i), a)] \\ &= I_n(a) - \frac{1}{n} \sum_{i=1}^n [y_i \frac{\psi'}{\psi}(x(i), t_{s_0}) + (1 - y_i) \frac{-\psi'}{1-\psi}(x(i), t_{s_0})] \\ &+ \frac{1}{n} \sum_{i=1}^n [y_i - R\{x(i)\}] [\frac{\psi'}{\psi}(x(i), t_{s_0}) + \frac{-\psi'}{1-\psi}(x(i), t_{s_0})] \\ &+ \frac{1}{n} \sum_{i=1}^n R\{x(i)\} [\frac{\psi'}{\psi}(x(i), t_{s_0}) - \frac{\psi'}{\psi}(x(i), a)] \\ &+ [1 - R\{x(i)\}] [\frac{-\psi'}{1-\psi}(x(i), t_{s_0}) - \frac{-\psi'}{1-\psi}(x(i), a)] \end{aligned}$$

It follows from (8.12) and (8.13) that the first part

$$\begin{aligned} \sup_{a \in [A, B]} |I_n(a) - \frac{1}{n} \sum_{i=1}^n [y_i \frac{\psi'}{\psi}(x(i), t_{s_0}) + (1 - y_i) \frac{-\psi'}{1 - \psi}(x(i), t_{s_0})]| \\ \leq \sup_{a \in [A, B]} \frac{1}{n} \sum_{i=1}^n [y_i \epsilon + (1 - y_i) \epsilon] \\ = \epsilon \end{aligned}$$

Similarly, we have the third part

$$\begin{aligned} \sup_{a \in [a, B]} |\frac{1}{n} \sum_{i=1}^n R\{x(i)\} [\frac{\psi'}{\psi}(x(i), t_{s_0}) - \frac{\psi'}{\psi}(x(i), a)] \\ + [1 - R\{x(i)\}] [\frac{-\psi'}{1 - \psi}(x(i), t_{s_0}) - \frac{-\psi'}{1 - \psi}(x(i), a)]| \\ \leq \epsilon \end{aligned}$$

Denote the second part $\frac{1}{n} \sum_{i=1}^n [y_i - R\{x(i)\}] [\frac{\psi'}{\psi}(x(i), t_{s_0}) + \frac{-\psi'}{1 - \psi}(x(i), t_{s_0})]$ as I_{n_2} .

We would like to establish that I_{n_2} goes to 0 almost surely for each t_{s_0} . For fixed t_{s_0} , $\{nI_{n_2} : n \geq 1\}$ forms a martingale and the terms in the summation in I_{n_2} are bounded. I_{n_2} tends to 0 almost surely under the limit theorem for the martingales.

Since there is a finite number of partition points, it follows that,

$$\sup |I_n(a) - \tilde{I}_n(a)| \rightarrow 0 \tag{8.14}$$

almost surely.

The next step is to consider the finite interval $S_{\Delta_0} = [a_{(1)}, a_{(m+|\Delta_0|)}]$ where $a_{(1)} = \min(a_1, \dots, a_{m+|\Delta_0|})$ and $a_{(m+|\Delta_0|)} = \max(a_1, \dots, a_{m+|\Delta_0|})$. **Condition D3** and the convexity of the set S indicate that $S_{\Delta_0} \subset S$. Define $\pi_n(d_i) \in [0, 1]$ as the

frequency that the level d_i has been assigned to the first n experiments. Then

$\tilde{I}_n(a)$ can be rewritten as

$$\tilde{I}_n(a) = \sum_{i=1}^{m+|\Delta_0|} \pi_n(d_i) \left[R(d_i) \frac{\psi'}{\psi}(d_i, a) + (1 - R(d_i)) \frac{-\psi'}{1 - \psi}(d_i, a) \right] \quad (8.15)$$

Let \tilde{a}_n be the solution to the equation

$$\tilde{I}_n(a) = \sum_{i=1}^{m+|\Delta_0|} \pi_n(d_i) \left[R(d_i) \frac{\psi'}{\psi}(d_i, a) + (1 - R(d_i)) \frac{-\psi'}{1 - \psi}(d_i, a) \right] = 0 \quad (8.16)$$

For each $1 \leq i \leq m + |\Delta_0|$, based on the definition of a_i and *Condition M2*, a_i is the unique solution of

$$R(d_i) \frac{\psi'}{\psi}(d_i, a) + (1 - R(d_i)) \frac{-\psi'}{1 - \psi}(d_i, a) = 0$$

It follows that \tilde{a}_n falls into the interval S_{Δ_0} . Since \hat{a}_n is the solution of $I_n(a) = 0$, (8.14) and uniform continuity ensure that, almost surely, $\hat{a}_n \in S$ for n sufficiently large, which leads to

$$|\psi(d_0, \hat{a}_n) - \theta| < |\psi(d_i, \hat{a}_n) - \theta|,$$

for $i = 1, \dots, m, d_i \neq d_0$ and d_0 is the true MTD for the "new" one-group study reduced from the two-group study. Thus, for n large enough $x(n+1) \equiv d_0$ or $x_1(n+1) \equiv x_{10}$ and $x_2(n+1) \equiv x_{20}$.

As $n \rightarrow \infty$, $\pi_n(d_i)$, ($i = 1, \dots, m + |\Delta_0|$) in (8.15) become negligible except

$\pi_n(d_0)$, which tends to 1. As the solution to (8.16), \tilde{a}_n will tend to the solution to

$$R(d_0) \frac{\psi'}{\psi}(d_0, a) + (1 - R(d_0)) \frac{-\psi'}{1 - \psi}(d_0, a) = 0$$

Therefore, as $n \rightarrow \infty$, $\tilde{a}_n \rightarrow a_0$ in probability. Because of (8.14), as $n \rightarrow \infty$, $\hat{a}_n \rightarrow a_0$ in probability. As a consequence, as $n \rightarrow \infty$, $x_1(n_1 + 1) \rightarrow x_{10}$ and $x_2(n_2 + 1) \rightarrow x_{20}$ in probability.

Chapter 9

Summary and Remarks

There has been some concern whether a class of one-parameter models is rich enough to describe the dose-toxicity relationship. We need to remember that the overall fit of the model is not the primary goal of a phase I oncology study. The aim of a phase I study is to locate the optimal dose, MTD, and also to allocate as many patients as possible to MTD. Although it seems natural to use the logistic model, the nature of dose-finding studies makes it less than optimal. The logistic model is too flexible and the scarcity of data from a phase I trial, especially at the early stage, causes a wide range of parameter estimates. The dose-toxicity curve from the logistic model is often non-informative except at or close to the MTD. The logistic model can even have a zero, negative or infinite estimate of the slope, which could recommend the subsequent patient to a dangerous dose. Our simulation has shown that the logistic model put more early patients to high untested doses, compared with the two-stage one-parameter CRM. The two-parameter CRM is incoherent and more aggressive. While the Bayesian prior can provide some stability to the logistic model, there remains instability and it takes longer than we

might imagine for things to stabilize.

The one-parameter CRM can either take on a two-stage design with the 1st stage as an up-down design, or a one-stage design with a prior for its parameter as in the two-parameter CRM. The advantage of the two-stage one-parameter CRM is that it has no need for Bayesian prior distributions. The influence of priors could be dramatic. It seems ideal that we can incorporate any known information into the model. However, the wrong specification of prior information can hamper attempts to find the correct MTD and also put more patients to less effective low doses and/or more toxic high doses. The two-parameter CRM needs priors to start. It cannot have a two-stage design as in the one-parameter CRM. Specification of formal distributions for parameters of logistic model causes tremendous computation difficulty and burden. One option is to use pseudo-data, the imaginary patients. The popular pseudo-data used in ADEPT and some articles put the first patient to the lowest dose. Our simulation has shown the two-parameter CRM performs poorly when the treatment is very safe, compared with the two-stage one-parameter CRM. The two-parameter CRM with pseudo data has a much lower chance to find the right MTD, especially when the number of doses is small. When there are only 3 doses, the performance of the two-parameter model is less than half of the one-parameter one. As a direct result, the two-parameter CRM assigns many more patients to the less effective low doses.

In reality we are always dealing with patient heterogeneity since no two patients are identical. As in other types of clinical trials we are essentially looking for an average effect. However, we may sometimes be in a position to specifically address the issue of patient heterogeneity as in the motivating example for the shift model. If certain prognostic factors are known to significantly affect

patients' treatment, it makes sense to learn the separate treatment information for each prognostic group instead of the average information. To our intuition and also demonstrated by this thesis, ignoring the group difference, the recommended average MTD is toxic to one group while it is sub-optimal to another group. The designs that acknowledge the group difference (CRM shift model, CRM two-parameter two-group model and two separate CRM designs) all outperform the design ignoring population heterogeneity. The two-separate-study design fails to utilize the information common to both groups. Furthermore in the absence of population heterogeneity, the performance of the CRM shift model is nearly the same as the combined one-group CRM design. The performance of the CRM shift model is nearly optimal as well. With incorporation of the prior information for the levels shifted, the performance of the CRM shift model can be further enhanced. It is straightforward to specify this prior for the levels shifted comparing the prior specification in the two-group two-parameter model.

Besides the issues and methods to stop a single group study, an additional but unique issue in the two-group studies is whether the two groups should stop recruitment simultaneously. Clearly stopping both groups simultaneously when both groups meet the stopping rule (such as the $m + 1$ rule) requires more patients and, as a consequence, has better performance. However the simulation shows stopping simultaneously will prolong the study 32% while the accuracy gain in finding the right MTDs is only 5%, compared with stopping at different times. Therefore, stopping separately is preferred in order to quickly proceed to the phase II trials, which is especially true for often slowly accruing oncology trials.

Like all other statistical analysis involving with categorical variables, different

classification or misclassification of a multi-level prognostic factor or a continuum into two groups will affect the practical utility of the two-group CRM model. There is no surprise that classification has little effect on model performance when the effect of the classified factor is weak. The effect of classification is stronger when the effect of a prognostic factor on patient's sensitivity to treatment is strong. Different classifications will produce different toxicity rates for the MTDs and even different MTD locations. Overall, classifying a heterogeneous population into two less heterogeneous groups leads to more accurate MTD estimation. Clearly, elevation of misclassification will dampen the model performance. Misclassification causes each group to be less homogenous within the group and the group difference is weakened. Therefore, the group effect on the MTD is attenuated. The performance of the two-group CRM shift model with misclassified patients will be between the one-group CRM model ignoring group difference and the two-group CRM shift model with accurate classification.

In practice, we are inevitably faced with small-to-moderate sample sizes. However the result of asymptotic consistency will be helpful in that it can not only provide some assurance as to the basic statistical soundness of the model but can also provide guidance in the choice and structure of different candidate working models. If a suggested method fails to converge asymptotically, then it seems likely that it will perform poorly in the small sample case. We have shown consistency of the two-group CRM shift model, which further strengthens the usability of the model.

This thesis demonstrated how CRM designs can be extended to two or multiple populations. For most instances, we should not use the two separate study design or the combined one group design. The CRM shift model and the two-

parameter model are particularly developed for heterogeneous populations. We especially recommend the CRM shift model. On average, it has better performance. It is easier to visualize and quantify the magnitude of heterogeneity in the CRM shift model. Both Bayesian framework and likelihood framework can be utilized in the CRM shift model. As shown in Section 4.2, we can specify the prior distribution of the number of doses shifted, Δ , where $\Delta \in \{-k + 1, -k + 2, \dots, k - 2, k - 1\}$. In the simulated Table A.16, the prior in the first example is $(0, 0, 0, 0, 0, 0.33, 0.33, 0.33, 0, 0, 0)$ for $\Delta = (-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)$. It is much more complicated to specify priors for the 2-group 2-parameter CRM model. The relationship between toxicity and parameters is not of a linear function. The representation of the amount of heterogeneity between groups depends not only on parameter b , but also on parameter a , or the probability curve for the first group. For example, let $f(y)$ and $f(z)$ as the true distributions of parameters a and b . In reality, for example, the specified prior distribution $f_*(y)$ significantly deviates from the true one. Although the prior probability of b is perfectly close to $f(z)$, $\psi(x, f_*(y), f(z))$ could be completely different from $\psi(x, f(y), f(z))$. Only when both priors are specified correctly, the extra information from priors will improve the study performance. In the CRM shift model, even if the prior distribution of parameter a is misspecified, the group difference which is reflected by the number of doses shifted can be specified accurately. It is due to the independent relation between G_1 MTD and Δ . In other words, the number of doses shifted does not depend on the location of G_1 MTD.

Appendix A

Tables

Table A.1: Summary of the four dose-escalation designs and the two intra-patient dose escalation options for the accelerated titration design

Design		Description
1	3 + 3	Cohort of 3 patients per dose level. If 1/3 patients on dose i has DLT, add additional 3 patients. If 0/3 patients has DLT, escalate to dose $i + 1$. Otherwise, dose $i - 1$ is the MTD.
2	ATD-1	Cohorts of 1 patient per dose level. When 1st episode of 1st-course DLT is observed, or second instance of 1st-course grade 2 toxicity of any type, expand cohort for current dose level and revert to design 1 for all future cohorts
3	ATD-2	Same as design 2, except that double-dose steps are used during initial accelerated stage of trial
4	ATD-3	Cohort of 1 per dose level and double-dose steps are used during the initial accelerated stage of the trial. When the 1st instance observed at any course, or the 2nd instance of any course grade 2 toxicity of DLT is of any type, expand cohort for current dose level and revert to design 1 for all future cohorts. When the first instance of moderate toxicity is observed, two additional patients must have been treated at that dose, or a higher dose, without experiencing moderate or worse toxicity, in order that the accelerated phase continues
Escalation		Description
A		No intra-patient dose escalation. Deescalate if DLT at previous course
B		Escalate if grade 0-1 toxicity at previous course. Deescalate if DLT at previous course

Table A.2: Example of study outcomes from the one-group CRM design

Patient _{<i>j</i>}	Dose (X_j)	DLT(Y_j)
1	1	0
2	2	0
3	3	0
4	4	1
5	2	0
6	3	0
7	3	1
8	2	0
9	2	0
10	3	0
11	3	0
12	3	0
13	3	0
14	3	1
15	3	0
16	3	0
.	.	.
.	.	.
.	.	.
n	.	.

Table A.3: Example of Predicted Toxicity Rates

d_i	1	2	3	4	5	6
α_i	0.1	0.2	0.3	0.4	0.5	0.6
$\hat{R}(d_i)$	0.045	0.115	0.198	0.292	0.394	0.503

The working codes are 0.10, 0.20, 0.30, 0.40, 0.50, and 0.60 for doses 1 to 6

Table A.4: An example of patient assignment based on the two-parameter CRM

Patient	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Dose	1	6	5	6	6	6	1	6	6	6	6	6	6	6	6	6
Toxicity	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1

The true underlying toxicity rates are 0.35, 0.45, 0.55, 0.70, 0.80 and 0.95 for doses 0.6, 1.2, 2.0, 3.0, 4.0, and 5.0 mg

Table A.5: The toxicity rate of six scenarios to demonstrate the performance of CRM

Scenario	Dose					
	1	2	3	4	5	6
1	0.35	0.45	0.55	0.70	0.80	0.95
2	0.25	0.35	0.45	0.55	0.70	0.80
3	0.15	0.25	0.35	0.45	0.55	0.70
4	0.10	0.15	0.25	0.35	0.45	0.55
5	0.05	0.10	0.15	0.25	0.35	0.45
6	0.02	0.05	0.10	0.15	0.25	0.35

For 2-parameter CRM we take the real doses as 0.6, 1.2, 2, 3, 4, 5 (mg). For one-parameter CRM we use the power model with $\alpha_i, i = 1, \dots, 6$ as 0.10, 0.20, 0.30, 0.40, 0.50, 0.60.

Table A.6: Prior Information Expressed by Pseudo-data

Dose	Number of Patients	Number of DLT
d_{-1}	n_{-1}	t_{-1}
d_0	n_0	t_0

Table A.7: Default Setting for Pseudo-data

Dose	Number of Patients	Number of DLT
d_1	3	3θ
d_k	3	1.5

Table A.8: The toxicity rates of very safe scenarios (the true MTD is the last dose)

Scenario	Dose						
	1	2	3	4	5	6	7
3	0.05	0.10	0.22				
4	0.05	0.10	0.15	0.22			
5	0.01	0.05	0.10	0.15	0.22		
6	0.01	0.04	0.08	0.12	0.16	0.22	
7	0.01	0.04	0.07	0.11	0.14	0.17	0.22

For 2-parameter CRM we take the real doses as 0.20, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75 (mg). For one-parameter CRM we use the power model with $\alpha_i, i = 1, \dots, 7$ as 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70.

Table A.9: Pseudo-data Used for Simulations in Section 2.2.2

Dose	Number of Patients	Number of DLT
d_1	1	0.2
d_k	1	0.5

Table A.10: Scenarios of true toxicity rates for the 2-group CRM shift model

Scenario	Group	Probability of toxicity $R(d_i)$					
A	1	.07	.23	.31	.35	.45	.57
	2	.07	.23	.31	.35	.45	.57
B	1	.08	.20	.35	.50	.70	.80
	2	.01	.05	.18	.40	.55	.70
C	1	.02	.19	.31	.45	.51	.63
	2	.03	.05	.11	.21	.39	.50
D	1	.7	.23	.31	.35	.45	.57
	2	.01	.03	.05	.09	.20	.40

Table A.11: α_i for the power model used in the simulations

Codes	α_i					
	d_1	d_2	d_3	d_4	d_5	d_6
0	0.20	0.30	0.50	0.70	0.80	0.90
1	0.10	0.20	0.30	0.50	0.70	0.80
2	0.05	0.10	0.20	0.30	0.50	0.70

Table A.12: MTD recommendation and in-trial allocation for scenario A

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.07	.23	.31	.35	.45	.57	.07	.23	.31	.35	.45	.57
Ratio of $N_1/N_2 = .50/.50$												
Prop (MTD)												
Scheme I	.27	.49	.19	.04	.00	.00	.12	.45	.28	.13	.02	.00
Scheme II	.27	.47	.19	.05	.00	.00	.09	.40	.29	.16	.04	.00
Scheme III	.22	.38	.23	.11	.04	.01	.22	.39	.22	.12	.04	.01
Scheme IV	.17	.51	.23	.09	.01	.00	.17	.51	.23	.09	.01	.00
Prop(Pat)												
Scheme I	.35	.38	.19	.06	.02	.00	.19	.33	.25	.15	.06	.02
Scheme II	.35	.33	.19	.07	.03	.02	.17	.29	.26	.15	.08	.03
Scheme III	.31	.29	.20	.11	.06	.03	.31	.29	.20	.11	.06	.03
Scheme IV	.25	.37	.22	.11	.04	.01	.25	.37	.22	.11	.04	.01
Ratio of $N_1/N_2 = .75/.25$												
Prop (MTD)												
Scheme I	.21	.50	.21	.07	.01	.00	.11	.35	.28	.19	.06	.01
Scheme II	.23	.50	.20	.06	.01	.00	.09	.32	.28	.18	.10	.03
Scheme III	.19	.46	.21	.11	.02	.00	.28	.27	.22	.16	.05	.03
Scheme IV	.17	.51	.23	.09	.01	.00	.17	.51	.23	.09	.01	.00
Prop (Pat)												
Scheme I	.29	.37	.20	.09	.03	.01	.17	.26	.23	.20	.09	.04
Scheme II	.31	.36	.19	.09	.03	.01	.14	.25	.24	.18	.12	.07
Scheme III	.28	.33	.20	.12	.06	.02	.38	.25	.17	.10	.07	.04
Scheme IV	.25	.37	.22	.11	.04	.01	.25	.37	.22	.11	.04	.01

Table A.13: MTD Recommendation and in-trial allocation for scenario B

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.08	.20	.35	.50	.70	.80	.01	.05	.18	.40	.55	.70
Ratio of $N_1/N_2 = .50/.50$												
Prop (MTD)												
Scheme I	.18	.54	.27	.01	.00	.00	.00	.19	.61	.19	.01	.00
Scheme II	.23	.49	.26	.01	.00	.00	.00	.12	.63	.21	.03	.00
Scheme III	.22	.47	.26	.04	.00	.00	.00	.15	.62	.21	.02	.00
Scheme IV	.02	.43	.53	.03	.00	.00	.02	.43	.53	.03	.00	.00
Prop(Pat)												
Scheme I	.24	.40	.29	.06	.01	.00	.07	.22	.43	.22	.05	.01
Scheme III	.32	.35	.25	.05	.02	.01	.06	.17	.44	.23	.08	.03
Scheme II	.30	.33	.24	.09	.03	.01	.10	.20	.41	.20	.07	.02
Scheme IV	.11	.35	.42	.09	.02	.01	.11	.35	.42	.09	.02	.01
Ratio of $N_1/N_2 = .75/.25$												
Prop (MTD)												
Scheme I	.16	.62	.22	.01	.00	.00	.02	.26	.46	.23	.03	.00
Scheme II	.19	.60	.20	.01	.00	.00	.01	.18	.48	.25	.07	.02
Scheme III	.19	.58	.22	.01	.00	.00	.03	.19	.41	.30	.05	.02
Scheme IV	.07	.59	.34	.01	.00	.00	.07	.59	.34	.01	.00	.00
Prop(Pat)												
Scheme I	.24	.45	.26	.05	.01	.00	.10	.24	.33	.23	.08	.02
Scheme III	.28	.42	.23	.05	.01	.00	.07	.19	.35	.23	.12	.05
Scheme II	.28	.39	.24	.06	.02	.00	.19	.23	.29	.17	.09	.04
Scheme IV	.17	.42	.32	.07	.02	.00	.17	.42	.32	.07	.02	.00

Table A.14: MTD Recommendation and in-trial allocation for scenario C

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.02	.19	.31	.45	.51	.63	.03	.05	.11	.21	.39	.50
Ratio of $N_1/N_2 = .50/.50$												
Prop (MTD)												
Scheme I	.07	.47	.39	.07	.01	.00	.00	.07	.32	.48	.12	.01
Scheme II	.11	.47	.32	.08	.02	.00	.01	.06	.32	.38	.20	.03
Scheme III	.11	.46	.34	.07	.01	.00	.00	.04	.30	.43	.20	.03
Scheme IV	.01	.23	.55	.19	.02	.00	.01	.23	.55	.19	.02	.00
Prop(Pat)												
Scheme I	.16	.38	.34	.09	.03	.00	.05	.13	.29	.35	.15	.03
Scheme II	.23	.34	.28	.10	.03	.02	.06	.11	.27	.30	.18	.08
Scheme II	.24	.34	.25	.10	.05	.02	.10	.14	.26	.26	.16	.08
Scheme IV	.08	.24	.39	.20	.06	.02	.08	.24	.39	.20	.06	.02
Ratio of $N_1/N_2 = .75/.25$												
Prop (MTD)												
Scheme I	.06	.54	.36	.04	.00	.00	.00	.11	.30	.43	.14	.01
Scheme II	.09	.56	.31	.04	.00	.00	.00	.07	.27	.33	.22	.10
Scheme III	.08	.55	.31	.06	.01	.00	.03	.11	.25	.38	.13	.10
Scheme IV	.02	.42	.47	.08	.01	.00	.02	.42	.47	.08	.01	.00
Prop(Pat)												
Scheme I	.15	.41	.32	.08	.03	.01	.07	.15	.25	.33	.17	.04
Scheme II	.19	.40	.28	.09	.03	.01	.05	.13	.24	.26	.19	.13
Scheme II	.18	.39	.28	.09	.04	.01	.20	.18	.22	.19	.13	.09
Scheme IV	.11	.34	.36	.13	.04	.01	.11	.34	.36	.13	.04	.01

Table A.15: MTD Recommendation and in-trial allocation for scenario D

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.07	.23	.31	.35	.45	.57	.01	.03	.05	.09	.20	.40
Ratio of $N_1/N_2 = .50/.50$												
Prop (MTD)												
Scheme I	.05	.42	.36	.14	.03	.00	.00	.01	.09	.47	.37	.06
Scheme II	.21	.39	.23	.13	.03	.00	.00	.01	.06	.23	.52	.18
Scheme III	.21	.38	.22	.13	.05	.01	.00	.01	.08	.22	.48	.20
Scheme IV	.01	.18	.34	.32	.14	.01	.01	.18	.34	.32	.14	.01
Prop(Pat)												
Scheme I	.17	.34	.29	.12	.06	.01	.05	.08	.17	.35	.27	.08
Scheme II	.31	.29	.20	.12	.06	.02	.04	.07	.12	.22	.33	.22
Scheme III	.30	.29	.20	.12	.06	.03	.08	.09	.13	.20	.29	.21
Scheme IV	.09	.20	.26	.24	.15	.05	.09	.20	.26	.24	.15	.05
Ratio of $N_1/N_2 = .75/.25$												
Prop (MTD)												
Scheme I	.10	.47	.29	.11	.02	.00	.00	.03	.17	.48	.25	.06
Scheme II	.20	.47	.22	.10	.02	.00	.00	.02	.09	.24	.36	.29
Scheme III	.19	.46	.21	.11	.02	.00	.00	.05	.10	.33	.29	.23
Scheme IV	.06	.38	.33	.19	.05	.00	.06	.38	.33	.19	.05	.00
Prop(Pat)												
Scheme I	.20	.36	.25	.12	.05	.02	.07	.10	.20	.20	.34	.09
Scheme II	.28	.34	.21	.11	.05	.02	.05	.09	.15	.25	.21	.24
Scheme III	.28	.33	.20	.12	.06	.02	.16	.15	.17	.18	.18	.16
Scheme IV	.16	.30	.26	.17	.08	.03	.16	.30	.26	.17	.08	.03

Table A.16: Simulation results for the shift model with prior probabilities of the levels shifted

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.08	.20	.35	.50	.70	.80	.01	.05	.18	.40	.55	.70
	$prior^b = 0.33, 0.33, 0.33$											
Prop (MTD)	.18	.54	.27	.01	.00	.00	.00	.19	.61	.19	.01	.00
Prop (Pat)	.24	.40	.29	.06	.01	.00	.07	.22	.43	.22	.05	.01
	$prior = .25, .50, .25$											
Prop (MTD)	.10	.56	.32	.01	.00	.00	.00	.12	.66	.21	.01	.00
Prop (Pat)	.22	.44	.29	.04	.01	.00	.06	.19	.48	.22	.04	.01
	$prior = .29, .43, .29$											
Prop (MTD)	.11	.57	.31	.01	.00	.00	.00	.14	.65	.20	.01	.00
Prop (Pat)	.21	.43	.30	.04	.01	.00	.06	.21	.45	.22	.05	.01
	$prior = .17, .50, .33$											
Prop (MTD)	.12	.62	.25	.00	.00	.00	.00	.09	.68	.22	.01	.00
Prop (Pat)	.24	.47	.25	.04	.01	.00	.06	.18	.47	.24	.05	.01
	$prior = .33, .50, .17$											
Prop (MTD)	.09	.58	.33	.01	.00	.00	.00	.15	.65	.19	.01	.00
Prop (Pat)	.20	.44	.31	.04	.01	.00	.06	.22	.46	.22	.04	.00
	$prior = .20, .60, .20$											
Prop (MTD)	.10	.64	.26	.01	.00	.00	.00	.10	.68	.20	.01	.00
Prop (Pat)	.22	.47	.25	.04	.01	.00	.06	.19	.48	.24	.04	.00

^b $prior$: the prior probability for $\Delta=0, -1, \text{ and } -2$

Table A.17: Example of incomplete information

	Dose					
	d_1	d_2	d_3	d_4	d_5	d_6
Y_{hi}	*	*	*	1	1	1
Y_{ji}	0	0	*	*	*	*

Table A.18: Example of complete information

	Dose					
	d_1	d_2	d_3	d_4	d_5	d_6
Y_{ji}	0	0	0	1	1	1

Table A.19: Simulation results to compare the shift model and the optimal design for scenario A

d_i	Group 1						Group 2					
	1	2	3	4	5	6	1	2	3	4	5	6
R_i	.07	.23	.31	.35	.45	.57	.07	.23	.31	.35	.45	.57
$p_i(16)$.27	.49	.19	.04	.00	.00	.12	.45	.28	.13	.02	.00
$q_i(16)$.23	.50	.14	.10	.03	.00	.23	.50	.14	.10	.03	.00

Table A.20: Simulation results to compare the shift model and the optimal design for scenario B

d_i	Group 1						Group 2					
	1	2	3	4	5	6	1	2	3	4	5	6
R_i	.08	.20	.35	.50	.70	.80	.01	.05	.18	.40	.55	.70
$p_i(16)$.18	.54	.27	.01	.00	.00	.00	.19	.61	.19	.01	.00
$q_i(16)$.20	.55	.22	.03	.00	.00	.00	.10	.71	.18	.01	.00

Table A.21: Levels of MTD shifted between groups for the combinations of the two prognostic factors

Combinations*	LH	LC	HH	HC
LH		0, 1	0, 1	0, 1, 2
LC			0	0, 1
HH				0, 1

* LH = Light+Hormone, LC = Light+Chemo, HH = Heavy+Hormone, and HC = Heavy+Chemo

Table A.22: Scenario of true toxicity rates for the 3-group CRM shift model

Group	d_1	d_2	d_3	d_4	d_5	d_6
G_1 HC	0.10	0.21	0.40	0.60	0.70	0.80
G_2 LC or HH	0.05	0.10	0.22	0.42	0.53	0.70
G_3 LH	0.00	0.05	0.10	0.20	0.39	0.56

* LH = Light+Hormone, LC = Light+Chemo, HH = Heavy+Hormone, and HC = Heavy+Chemo

Table A.23: Performance comparison with equal patient recruitment ¹

Group	outcome		d_1	d_2	d_3	d_4	d_5	d_6	
G_1	R(d_i)		0.10	0.21	0.40	0.60	0.70	0.80	
	MTD	Shift	0.11	0.63	0.25	0.00	0.00	0.00	
		Shift with Prior ²	0.19	0.69	0.12	0.00	0.00	0.00	
		Separate Studies	0.30	0.49	0.20	0.01	0.00	0.00	
	Treatment	Shift	0.21	0.47	0.28	0.03	0.00	0.00	
		Shift with Prior ²	0.32	0.52	0.14	0.01	0.00	0.00	
		Separate Studies	0.36	0.35	0.21	0.06	0.02	0.00	
	G_2	R(d_i)		0.05	0.10	0.22	0.42	0.53	0.70
		MTD	Shift	0.00	0.29	0.63	0.08	0.00	0.00
Shift with Prior ²			0.00	0.23	0.70	0.07	0.00	0.00	
Separate Studies			0.05	0.27	0.51	0.14	0.02	0.00	
Treatment		Shift	0.07	0.29	0.48	0.14	0.02	0.00	
		Shift with Prior ²	0.07	0.25	0.52	0.14	0.01	0.00	
		Separate Studies	0.17	0.25	0.35	0.15	0.06	0.02	
G_3		R(d_i)		0.00	0.05	0.10	0.20	0.39	0.56
		MTD	Shift	0.00	0.05	0.42	0.49	0.04	0.00
	Shift with Prior ²		0.00	0.00	0.25	0.68	0.06	0.00	
	Separate Studies		0.00	0.04	0.27	0.48	0.19	0.02	
	Treatment	Shift	0.04	0.12	0.38	0.37	0.07	0.01	
		Shift with Prior ²	0.03	0.06	0.27	0.50	0.11	0.01	
		Separate Studies	0.09	0.12	0.24	0.30	0.18	0.07	

¹ Sample sizes are 16, 16, and 16 for G_1 , G_2 and G_3 , respectively.² The prior probabilities for $\Delta_1' = 1$ and $\Delta_2' = 1$ are both 0.80.

Table A.24: Performance comparison with unequal patient recruitment ¹

Group	outcome		d_1	d_2	d_3	d_4	d_5	d_6	
G_1	R(d_i)		0.10	0.21	0.40	0.60	0.70	0.80	
	MTD	Shift	0.09	0.61	0.29	0.01	0.00	0.00	
		Shift with Prior ²	0.19	0.68	0.13	0.00	0.00	0.00	
		Separate Studies	0.30	0.43	0.24	0.02	0.00	0.00	
	Treatment	Shift	0.20	0.45	0.30	0.03	0.00	0.00	
		Shift with Prior ²	0.33	0.52	0.14	0.01	0.00	0.00	
		Separate Studies	0.38	0.31	0.21	0.07	0.02	0.01	
	G_2	R(d_i)		0.05	0.10	0.22	0.42	0.53	0.70
		MTD	Shift	0.00	0.27	0.65	0.07	0.00	0.00
Shift with Prior ²			0.00	0.21	0.72	0.07	0.00	0.00	
Separate Studies			0.02	0.28	0.60	0.10	0.01	0.00	
Treatment		Shift	0.07	0.29	0.49	0.13	0.02	0.00	
		Shift with Prior ²	0.07	0.25	0.53	0.13	0.01	0.00	
		Separate Studies	0.12	0.26	0.41	0.15	0.04	0.01	
G_3		R(d_i)		0.00	0.05	0.10	0.20	0.39	0.56
		MTD	Shift	0.00	0.06	0.42	0.48	0.04	0.00
	Shift with Prior ²		0.00	0.00	0.24	0.70	0.06	0.00	
	Separate Studies		0.00	0.05	0.31	0.40	0.20	0.04	
	Treatment	Shift	0.05	0.13	0.37	0.37	0.07	0.01	
		Shift with Prior ²	0.03	0.06	0.26	0.52	0.11	0.01	
		Separate Studies	0.11	0.15	0.23	0.26	0.17	0.08	

¹ Sample sizes are 12, 24, and 12 for G_1 , G_2 and G_3 , respectively.

² The prior probabilities for $\Delta_1' = 1$ and $\Delta_2' = 1$ are both 0.80.

Table A.25: Number of patients recruited by the two stopping rules

	Group 1						Group 2					
	Min	Q1	Q2	Mean	Q3	Max	Min	Q1	Q2	Mean	Q3	Max
SR1	6	12	15	16.4	20	36	6	12	16	17.1	20	36
SR2	6	15	20	21.5	27	36	6	15	21	21.6	27	36

Table A.26: Proportions of the recommended MTDs and treatments by the two stopping rules

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.08	.20	.35	.50	.70	.80	.01	.05	.18	.40	.55	.70
MTD-SR1	.18	.51	.30	.01	.00	.00	.01	.18	.64	.16	.01	.00
MTD-SR2	.15	.57	.27	.01	.00	.00	.00	.16	.68	.14	.01	.00
Trt-SR1	.25	.43	.27	.05	.01	.00	.06	.22	.46	.21	.05	.01
Trt-SR2	.25	.46	.25	.04	.01	.00	.05	.20	.50	.20	.04	.01

Table A.27: Toxicity rates under different classifications for the two scenarios

Threshold	Grp	Weak effect ($\beta = .50$)						Strong effect ($\beta = 2.0$)					
		d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
.50	R_h	.01	.07	.19	.35	.54	.71	.03	.19	.43	.62	.78	.88
	R_l	.01	.05	.16	.29	.48	.65	.01	.08	.22	.38	.57	.73
.66	R_h	.01	.07	.20	.36	.55	.71	.03	.21	.47	.66	.81	.90
	R_l	.01	.06	.16	.30	.49	.66	.01	.09	.25	.42	.61	.75

Table A.28: Proportions of recommended MTDs and treatments – Weak effect of prognostic factor classified at threshold of .50

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.01	.07	.19	.35	.54	.71	.01	.05	.16	.29	.48	.65
MTD	.01	.19	.64	.17	.00	.00	.00	.06	.48	.39	.06	.00
Treatment	.09	.24	.46	.17	.03	.00	.04	.12	.36	.32	.12	.03

Table A.29: Proportions of recommended MTDs and treatments – Weak effect of prognostic factor classified at threshold of .66

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.01	.07	.20	.36	.55	.71	.01	.06	.16	.30	.49	.66
MTD	.01	.21	.64	.15	.00	.00	.00	.07	.50	.38	.06	.00
Treatment	.09	.25	.47	.16	.03	.00	.05	.13	.37	.31	.12	.02

Table A.30: Proportions of recommended MTDs and treatments – Strong effect of prognostic factor classified at threshold of .50

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.03	.19	.43	.62	.78	.88	.01	.08	.22	.38	.57	.73
MTD	.15	.64	.21	.00	.00	.00	.01	.30	.52	.17	.01	.00
Treatment	.22	.49	.25	.03	.00	.00	.07	.28	.39	.20	.04	.01

Table A.31: Proportions of recommended MTDs and treatments – Strong effect of prognostic factor classified at threshold of .66

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.03	.21	.47	.66	.81	.90	.01	.09	.25	.42	.61	.75
MTD	.18	.68	.14	.00	.00	.00	.01	.39	.48	.11	.00	.00
Treatment	.25	.51	.21	.02	.00	.00	.08	.34	.38	.16	.03	.00

Table A.32: Effect of misclassification–Scenario B

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.08	.20	.35	.50	.70	.80	.01	.05	.18	.40	.55	.70
0%	.18	.52	.29	.01	.00	.00	.00	.19	.61	.19	.01	.00
5%	.17	.51	.30	.01	.00	.00	.00	.21	.59	.19	.01	.00
10%	.15	.51	.33	.01	.00	.00	.00	.23	.59	.17	.01	.00
15%	.14	.51	.34	.01	.00	.00	.01	.25	.57	.16	.01	.00
20%	.13	.50	.36	.01	.00	.00	.01	.27	.56	.16	.01	.00
25%	.11	.50	.36	.02	.00	.00	.01	.28	.56	.14	.01	.00
30%	.10	.50	.37	.02	.00	.00	.01	.30	.55	.14	.01	.00
35%	.09	.50	.39	.02	.00	.00	.01	.32	.54	.13	.01	.00
40%	.08	.49	.41	.02	.00	.00	.01	.32	.55	.11	.01	.00
45%	.07	.49	.42	.02	.00	.00	.01	.34	.53	.12	.00	.00
50%	.06	.48	.43	.02	.00	.00	.01	.35	.53	.10	.01	.00

Table A.33: Effect of misclassification–Scenario C

	Group 1						Group 2					
	d_1	d_2	d_3	d_4	d_5	d_6	d_1	d_2	d_3	d_4	d_5	d_6
$R(d_i)$.02	.19	.31	.45	.51	.63	.03	.05	.11	.21	.39	.50
0%	.07	.48	.39	.06	.01	.00	.00	.06	.33	.48	.13	.01
5%	.06	.46	.40	.07	.01	.00	.00	.08	.35	.45	.12	.01
10%	.06	.43	.43	.07	.01	.00	.00	.09	.35	.44	.11	.01
15%	.06	.42	.42	.08	.01	.00	.00	.11	.37	.41	.10	.01
20%	.06	.41	.44	.08	.01	.00	.00	.12	.39	.38	.10	.01
25%	.05	.39	.46	.09	.01	.00	.00	.13	.41	.36	.09	.01
30%	.05	.38	.46	.10	.01	.00	.00	.14	.42	.34	.08	.01
35%	.04	.37	.47	.11	.01	.00	.00	.16	.44	.32	.07	.01
40%	.04	.36	.48	.11	.01	.00	.00	.18	.44	.30	.07	.01
45%	.03	.34	.49	.12	.01	.00	.00	.19	.46	.28	.06	.01
50%	.03	.33	.50	.13	.01	.00	.00	.19	.46	.28	.06	.00

Table A.34: Illustration of the 2-group question reduced to a “new” 1-group question (the true $\Delta = 2$)

Level	G_1			d_1	d_2	d_3	d_4	d_5	d_6
	G_2	d_1	d_2	d_3	d_4	d_5	d_6		
	“new”	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
$R(d_i)$	G_1			$R_1(d_1)$	$R_1(d_2)$	$R_1(d_3)$	$R_1(d_4)$	$R_1(d_5)$	$R_1(d_6)$
	G_2	$R_2(d_1)$	$R_2(d_2)$	$R_2(d_3)$	$R_2(d_4)$	$R_2(d_5)$	$R_2(d_6)$		
	“new”	$R(d_1)$	$R(d_2)$	$R(d_3)$	$R(d_4)$	$R(d_5)$	$R(d_6)$	$R(d_7)$	$R(d_8)$
α	G_1			α_3	α_4	α_5	α_6	α_7	α_8
	G_2	α_1	α_2	α_3	α_4	α_5	α_6		
	“new”	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8

Appendix B

Figures

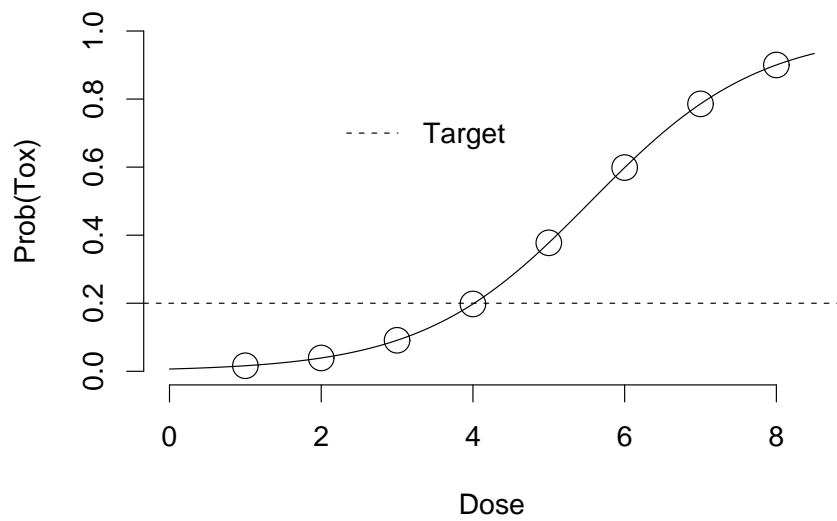


Figure B.1: Dose-toxicity profile

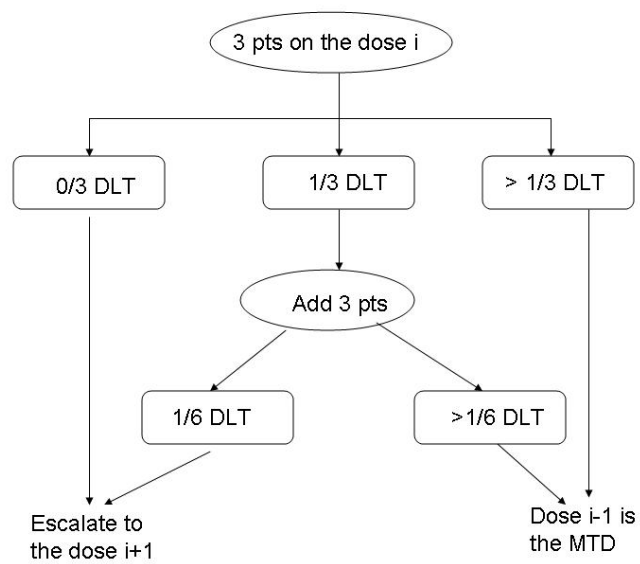


Figure B.2: Dose escalation scheme for the traditional 3 + 3 design

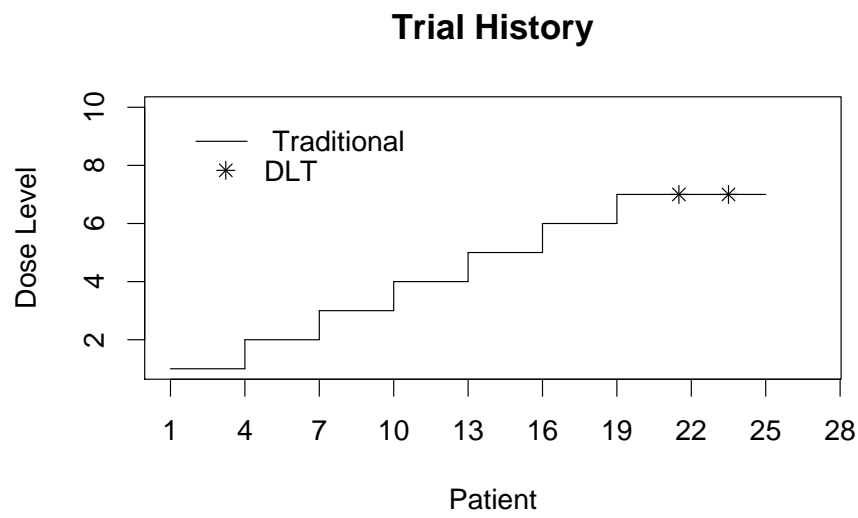
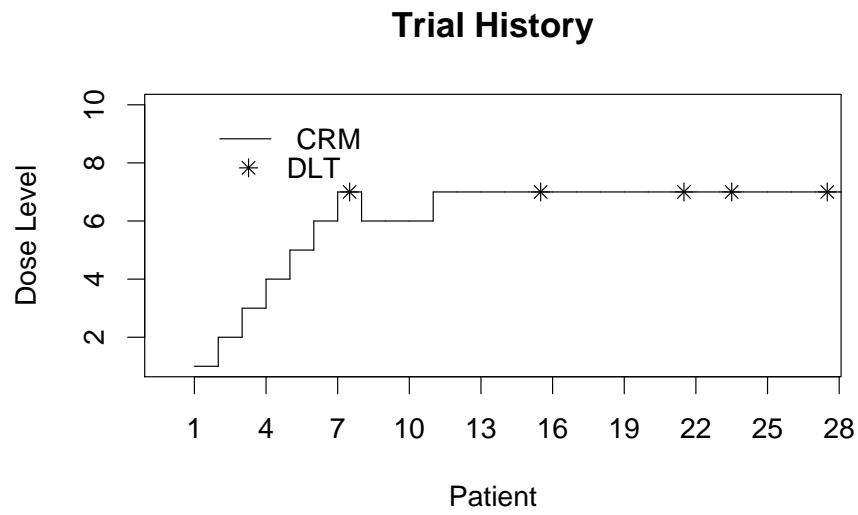


Figure B.3: Typical trial histories: CRM (top) and standard design (bottom)

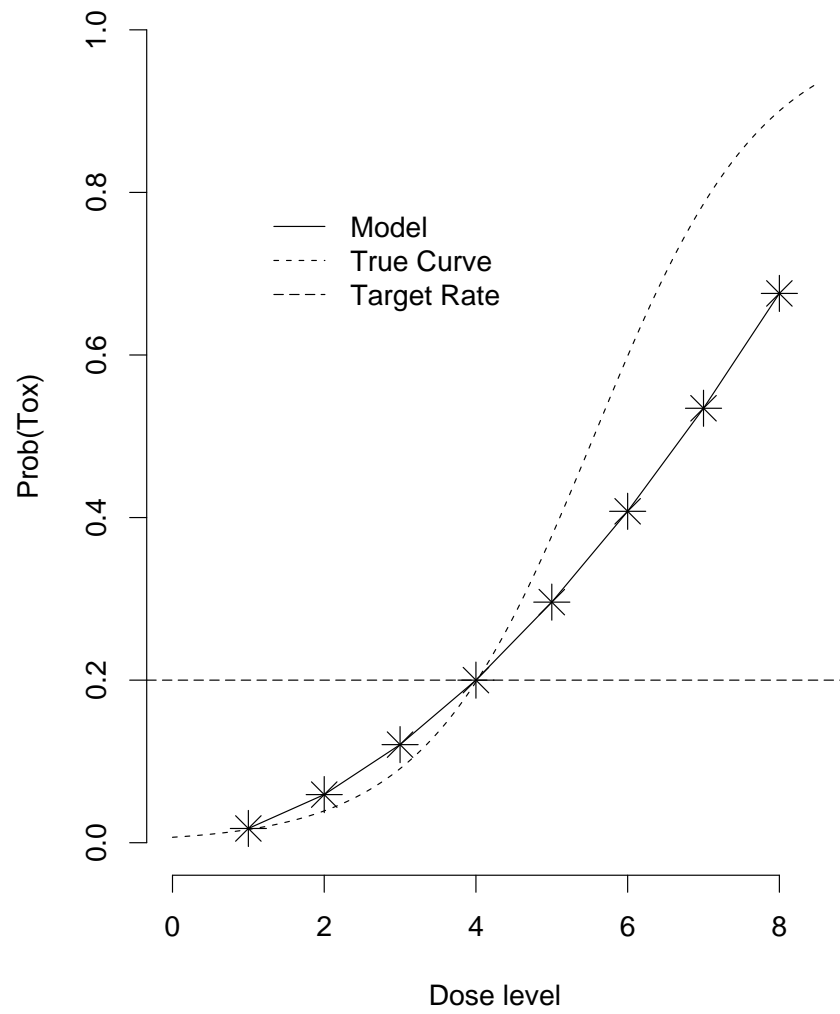


Figure B.4: Fit of the CRM working model

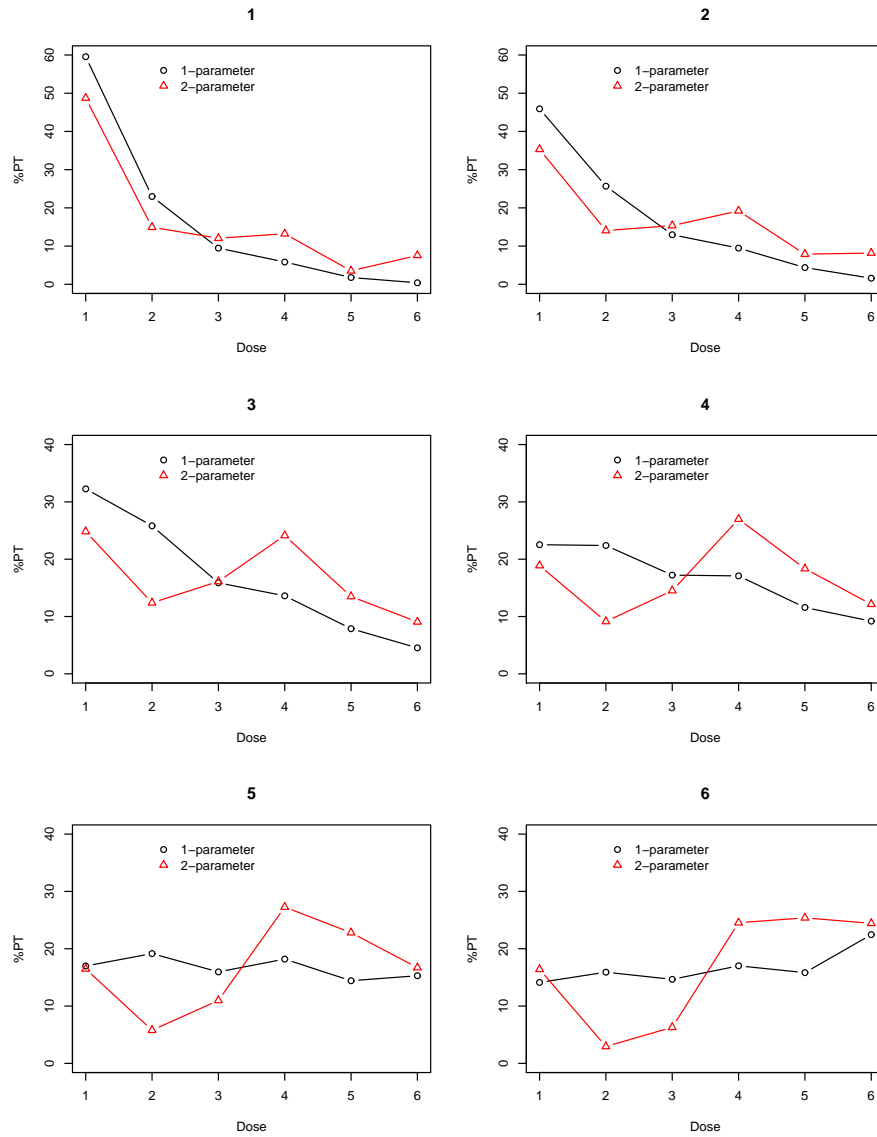


Figure B.5: The dose assigned to the first 8 patients. 1=True MTD at the first dose; 2=True MTD at the second dose; 3=True MTD at the third dose; 4=True MTD at the fourth dose; 5=True MTD at the fifth dose; 6=True MTD at the sixth dose

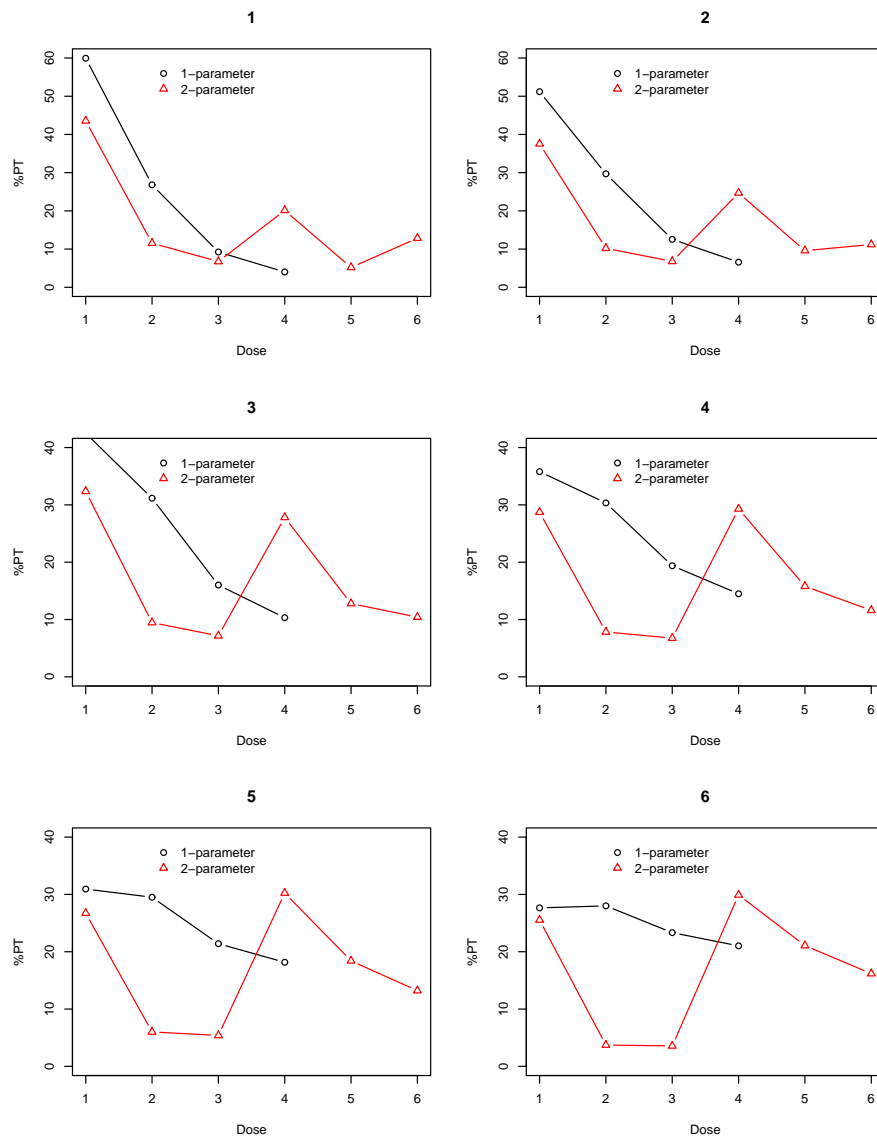


Figure B.6: The dose assigned to the first 4 patients. 1=True MTD at the first dose; 2=True MTD at the second dose; 3=True MTD at the third dose; 4=True MTD at the fourth dose; 5=True MTD at the fifth dose; 6=True MTD at the sixth dose

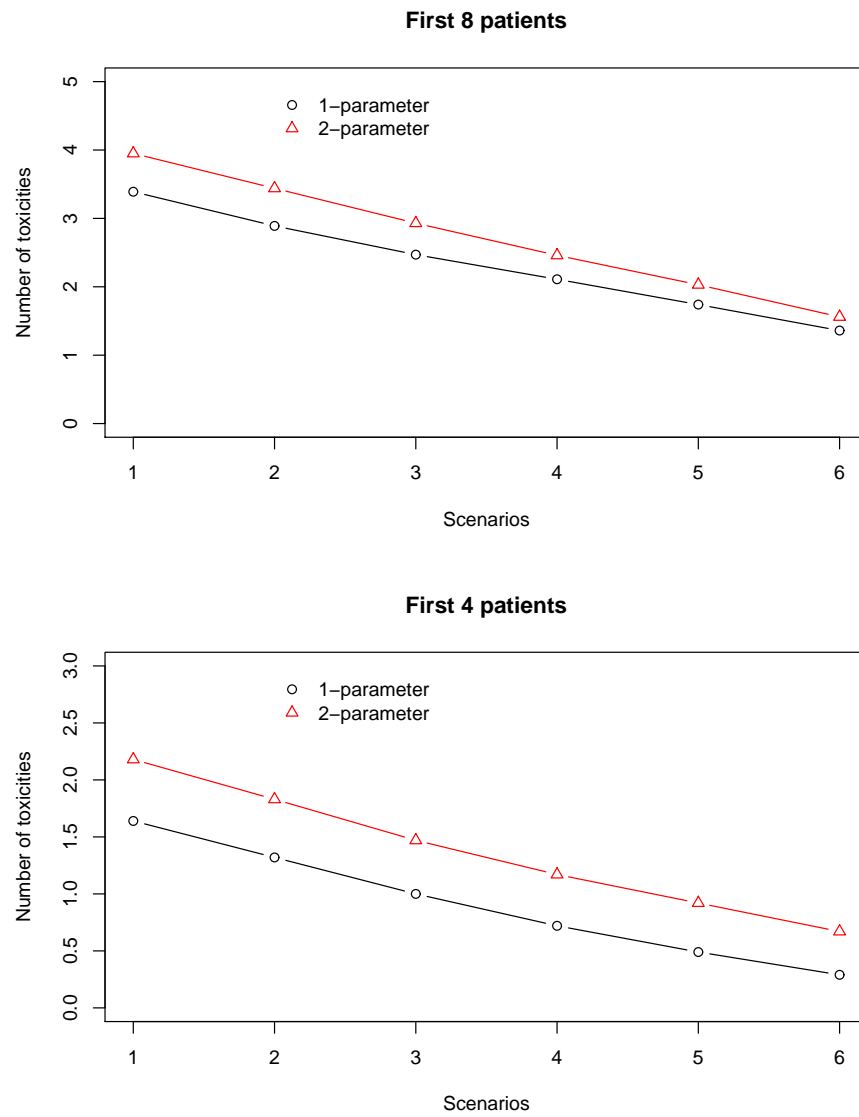


Figure B.7: Toxicities from the first 8 or 4 patients.

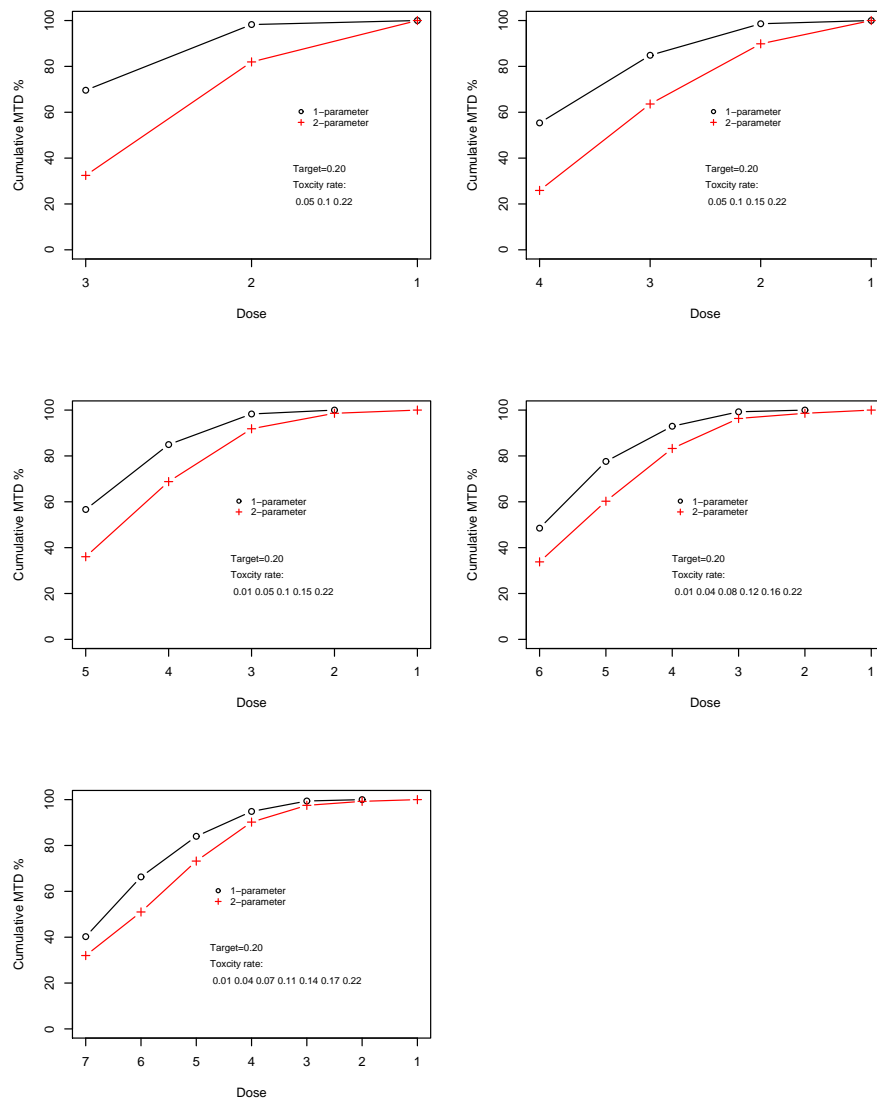


Figure B.8: The cumulative distribution of the estimated MTD.

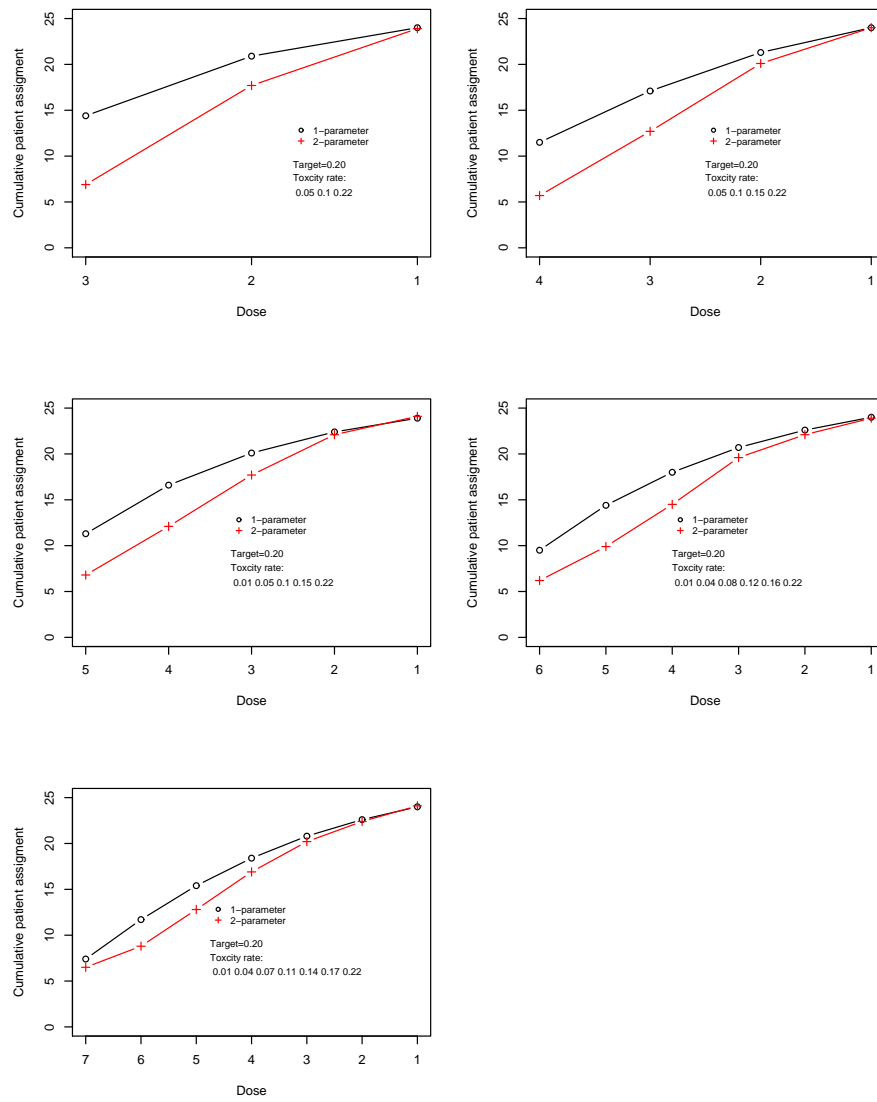


Figure B.9: Cumulative patient assignment.

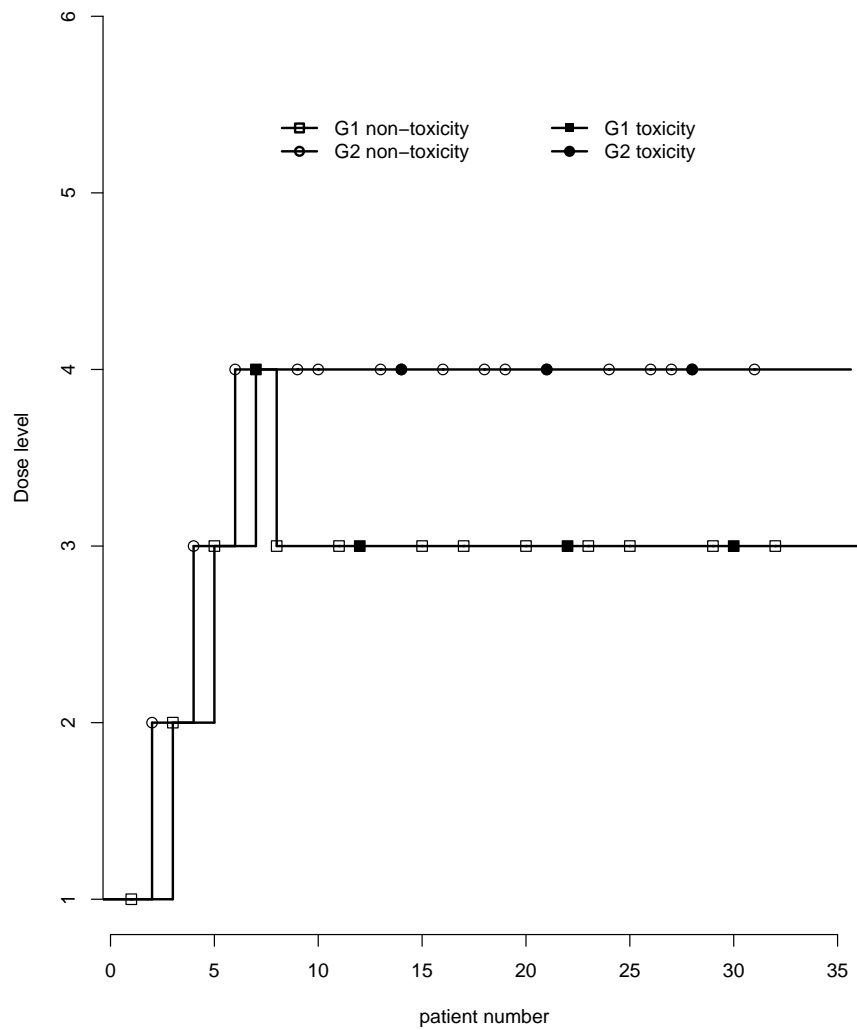


Figure B.10: Sample of trial history of the two-group CRM shift model

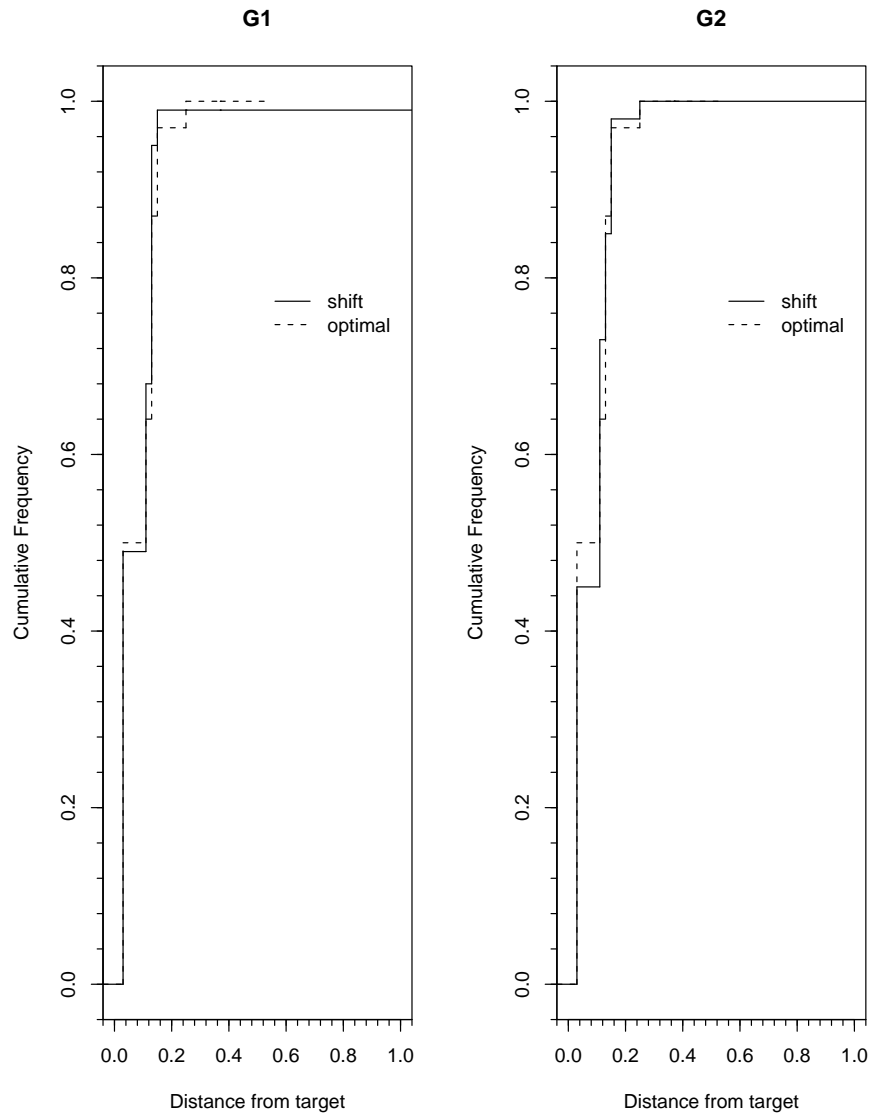


Figure B.11: Cumulative frequency plot of distance from target for scenario A

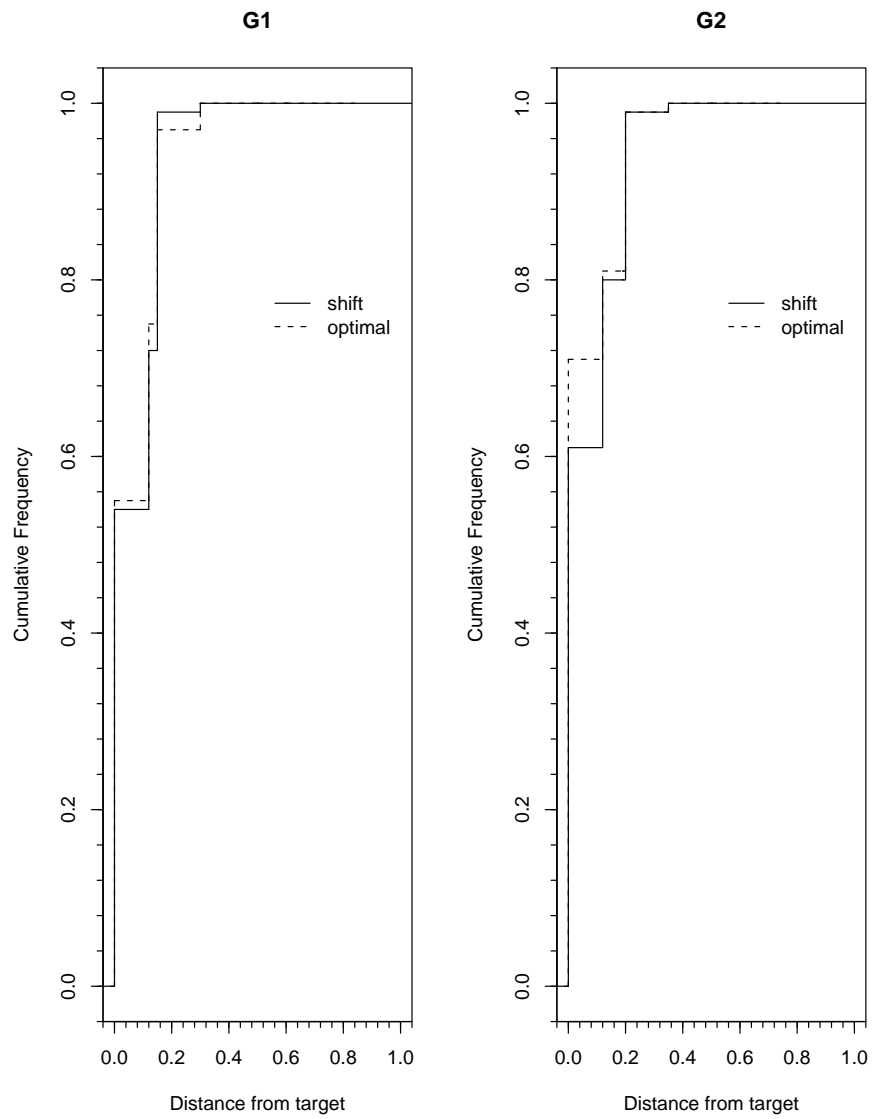


Figure B.12: Cumulative frequency plot of distance from target for scenario B

Bibliography

- [1] O'Quigley J., Pepe M. and Fisher L. (1990) Continual Reassessment Method: a practical design for Phase 1 clinical studies in cancer. *Biometrics* **46**, No 1, 33-48.
- [2] Carter SK. (1987). The phase I study, in *Fundamentals of Cancer Chemotherapy* (eds K.K, Hellmann and S.K. Carter), McGraw-Hill, New York, 285-300.
- [3] Edler L. (1990). Statistical requirements of phase I studies. *Onkologie* **13(2)**, 90-95.
- [4] Lin Y. and Shih WJ. (2004). Statistical properties of the modified algorithm-based designs for phase I cancer clinical trials. *Technical Report*, Department of Biostatistics, University of Medicine and Dentistry of New Jersey.
- [5] Storer EL. (1989). Design and analysis of phase I clinical trials. *Biometrics* **45**, 925-937.
- [6] Store EL. (1993). Small-sample confidence sets for the MTD in a phase I clinical trial. *Biometrics* **49**, 1117-1125.

- [7] Reiner E., Paoletti X., and O'Quigley J. (1999). Operating characteristics of the standard phase I clinical trial design. *Computational Statistics and Data Analysis* **30**, 303-315.
- [8] Iasonos A., Wilton AS., Riedel ER., Seshan VE., and Spriggs DR. (2008). A comprehensive comparison of the continual reassessment method to the standard 3 + 3 dose escalation scheme in Phase I dose-finding studies *Clin Trials* **5**(5), 465-477.
- [9] Sheiner LB. (1990). Implications of an alternative approach to dose-response trials. *J. Acquired Immune Deficiency Syndrome* **3**(suppl.2), 20-26.
- [10] Sheiner LB., Beal SL. and Sambol NC. (1989). Study designs for dose-ranging. *Clin. Pharmacology Theory* **46**(1), 63-77.
- [11] Sheiner LB., Hashimoto Y., and Beal SL. (1991). A simulation study comparing designs for dose ranging *Statistics in Medicine* **10**(3), 303-321.
- [12] Simon R., Freidlin B., Rubinstein L., Arbuch SG., Collins J., and Christian MC. (1997). Accelerated titration designs for phase I clinical trials in oncology. *J. Natl Cancer Inst.* **89**(15), 1138-1147.
- [13] Jordan SD., Poole CJ., Archer VR., Steven NM., and Burton A. (2003). retrospective evaluation of the feasibility of inpatient dose escalation as appropriate methodology for dose-ranging studies for combination cytotoxic regimens. *Cancer Chemother Pharmacol* **52**(2), 1113-1118.
- [14] Durham SD., and Flournoy N. (1994). Random walks for quantile estimation. *Statistical Decision Theory and Related Topics*, Gupta SS, Berger JO (eds). Springer: New York, 467-476.

- [15] Durham SD., Flournoy N., and Rosenberger WF. (1997). A random walk rule for phase I clinical trials. *Biometrics* **53**, 745-760.
- [16] Gezmu M., and Flournoy N. (2006). Group up-and-down designs for phase I clinical trials. *Journal of Statistical Planning and Inference* **136**, 1749-1764.
- [17] Collins JM , Grieshaber CK , and Chabner BA. (1990). Pharmacologically guided phase I clinical trials based upon preclinical drug development. *J Natl Cancer Inst* **82** (16), 1321-1326.
- [18] O'Quigley J, Hughes MD, Fenton T, Pei L. (2010) Dynamic calibration of pharmacokinetic parameters in dose-finding studies. *Biostatistics* **11** (3), 537-45.
- [19] Skolnik JM , Barrett JS , Jayaraman B , Patel D , and Adamson PC. (2008). Shortening the timeline of pediatric phase I trials: the rolling six design. *J Clin Oncol.* **26** (2), 1901-95.
- [20] Hartford S, Volchenbom SL, and Cohn SL. (2008). $3 + 3 \neq$ (Rolling) 6. *J Clin Oncol.* **26** (2), 1701-71.
- [21] Zhao L, Lee J, Mody R, Braun TM. (2011). The superiority of the time-to-event continual reassessment method to the rolling six design in pediatric oncology phase I trials. *Clinical Trials* **8**, 361-369.
- [22] O'Quigley J. and Reiner E. (1998) A stopping rule for the continual reassessment method. *Biometrika* **85**, 741-748.
- [23] Goodman SN, Zahurak ML and Piantadosi S. (1995). Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* **14**, 1149-1161.

- [24] Korn EL., Midthune D., Chen TT., Rubinstein LV., Christian MC., and Simon RM. (1994). A comparison of two phase I trial designs. *Statistics in medicine* **13**(18), 1799-1806.
- [25] O'Quigley J and Chevret S. (1991). Methods for dose finding studies of cancer clinical trials: a review and results of a Monte Carlo study. *Statistics in Medicine* **10**, 1647-1664.
- [26] Whitehead J. and Brunier H. (1995) Bayesian decision procedures for dose determining experiments. *Statistics in Medicine* **14**, 33-48.
- [27] Babb J., Rogatko A., and Zacks S. (1998). Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* **17**(10), 1103-1120.
- [28] Cheung YK , and Chappell R. (2000). Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* **56** (4), 1177-1182.
- [29] Normolle D , and Lawrence T. (2006). Designing dose-escalation trials with late-onset toxicities using the time-to-event continual reassessment method. *J Clin Oncol* **24** (27), 4426-4433.
- [30] O'Quigley J., Shen L. and Gamst A. (1999). Two-sample continual reassessment method. *J. Biopharm. Stat.* **9**, 17-44.
- [31] O'Quigley J. and Paoletti X. (2003). Continual Reassessment Method for Ordered Groups. *Biometrics* **59**, 430-440
- [32] O'Quigley J. (2006) Theoretical study of the continual reassessment method. *J. Statist. Planning. Inference* **136**, 1765-80.

- [33] Shen L. and O'Quigley J. (1996) Consistency of continual reassessment method under model misspecification. *Biometrika* **83**, 395-405.
- [34] O'Quigley J. and Shen L. Z. (1996). Continual reassessment method: A likelihood approach. *Biometrics* **52**, 163-174.
- [35] O'Quigley J. Paoletti X. and Maccario J. (2002) Non-parametric optimal design in dose finding studies. *Biostatistics* **3**, 51-56.
- [36] Cheung Y. K. (2005). Coherence principles in dose-finding studies. *Biometrika* **92**, 863-873
- [37] Bedding A. (2003). Bayesian ADEPT assisted decision making in early phase trials. *Pharmaceutical Statistics* **2**, 219-221.
- [38] Zhou Y. (2004). Choice of designs and doses for early phase trials. *Fundamental & Clinical Pharmacology* **18**, 373-378.
- [39] Chevret S. (2006) statistical methods for dose-finding experiments. ISBN-13: 978-0470861233
- [40] Gerke O. and Siedentop H. (2008). Optimal phase I dose-escalation trial designs in oncology –A simulation study. *Statistics in Medicine* **27**, 53295344.
- [41] Zhou Y. and Whitehead J. (2002). *Bayesian ADEPT: Operating Manual*. The University of Reading: Reading, U.K.
- [42] Paoletti X. and Kramar A. (2009). comparison of model choices for the continual reassessment method in phase I cancer trials. *Statistics in Medicine* **28**, 30123028.

- [43] Heyd J. and Carlin B. (1999). Adaptive design improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* **18**, 1307-1321.
- [44] O'Quigley J. (2002). Continual reassessment designs with early termination. *Biostatistics* **3**, 87-99.
- [45] Cheung Y.K. and Chappell R.(2002). A Simple technique to evaluate model sensitivity in the continual reassessment method. *Biometrics* **58**, 671-674.