DOSE-FINDING DESIGNS FOR PHASE II CLINICAL TRIALS

Changfu Xiao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics

Chapel Hill
2011

Approved by:

Anastasia Ivanova, PhD

Annelies Van Rie, PhD

Michael Hudgens, PhD

Shyamal Peddada, PhD

Bahjat Qaqish, PhD

ABSTRACT

CHANGFU XIAO: Dose-finding Designs for Phase II Clinical Trials (Under the direction of Anastasia Ivanova)

Most existing dose-finding designs have been proposed for Phase I oncology trials where the main outcome is toxicity, and dose escalation is guided by ethical considerations. Most of these designs have been developed assuming that dose-response curve is strictly increasing and the outcome is binary. The main outcome in Phase II non-oncology trials is efficacy, which is often continuous. A bivariate outcome combining efficacy and safety is also constantly considered in Phase II non-oncology trials. The goal of this work is to investigate the suitability of Phase I oncology designs for Phase II non-oncology trials and to develop dose-finding designs that better address the needs of Phase II non-oncology trials. Specifically, the first paper investigates which of the several known dose-finding methods is most suitable when dose response curve plateaus. Some of the designs tend to spread the allocation among the doses on the plateau, others, like the continual reassessment method and the *t*-statistic design, concentrate allocation at one of the doses with the *t*-statistic design selecting the lowest dose on the plateau more frequently. The second paper examines the optimal allocation for estimating the minimum effective and peak doses in a dose-ranging trial when the set of dose level is fixed and isotonic regression is used as a method of estimation. We propose fully sequential strategy for subject assignment. The proposed strategy includes adaptive randomization procedure to balance the allocation to placebo

and active doses. We also consider estimation in presence of covariates, where randomization procedure also balances the allocation with respect to a set of known covariates. Simulations show that the new adaptive strategy is superior to equal allocation. The third paper investigates a Bayesian adaptive two-stage design to efficiently estimate the minimum effective dose or the maximum dose in a dose-finding trial where some monotonicity assumptions regarding dose-response relationship can be made. The new design allocates subjects in stage 2 according to the posterior distribution of the location of the target dose. Simulations show that the proposed two-stage design is superior to equal allocation and to two-stage strategy where only one dose is left in stage 2.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION: ADAPTIVE DESIGN

In recent years, the cost of drug development has increased dramatically and the concerns for patient safety have also been escalating. Pharmaceutical companies and clinical researchers are facing greater pressure to reexamine traditional clinical trial techniques and improve the efficiency and safety of the clinical trial process. Adaptive design has emerged as one way to address those challenges. As the name suggests, it refers to "a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study" (United States Food and Drug Administration, 2010). Adaptive design methods may provide the same information with more efficiency and improved understanding of the treatment effect as compared to non-adaptive studies. The use of adaptive design methods in clinical trials has received significant attention from clinical scientists, biostatisticians, pharmaceutical companies, and regulatory agencies. For instance, in 2006, the Food and Drug Administration (FDA) (2006) released a Critical Path Opportunities List that calls for advancing innovative trial designs, especially for the use of prior experience or accumulated information in trial design. In 2010, the FDA released the draft guidance on Adaptive Design Clinical Trials for Drugs and Biologics identifying Phase I and Phase II studies as the most promising applications of adaptive designs.

Commonly considered adaptive design methods in clinical trials include, but are not limited to: adaptive dose-finding designs, adaptive randomization, response-adaptive designs,

adaptive seamless phase II/III trial designs, group sequential designs, and designs allowing sample size re-estimation. In this proposal, we will concentrate on adaptive designs for Phase II clinical trials.

Most of the existing dose-finding designs have been proposed for Phase I oncology trials where the main outcome is toxicity, and dose escalation is guided by ethical considerations. These designs are fully sequential, experimentation starts at the lowest dose and dose is gradually escalated. Most of these designs have been developed under the assumption that dose-response curve is strictly increasing and outcome is binary. The main outcome in Phase II non-oncology trials is efficacy, which is often continuous, or bivariate outcome, efficacy and safety, is often considered. Dose-response curve is assumed non-decreasing, as opposed to strictly increasing curve in Phase I trials, and there is no ethical constrain to start at the lowest dose. Some of the designs originally proposed for oncology trials have been used in Phase II non-oncology trials (e.g., Hall et al., 2005) without paying attention to the limitations described above. Only a small number of dose-finding designs have been proposed specifically for Phase II non-oncology trials (Berry et al., 2001; Ivanova et al., 2008; Ivanova et al., 2009; Dragalin and Fedorov, 2006; Miller et al., 2007).

The goal of this work is to investigate the suitability of Phase I oncology designs for Phase II non-oncology trials (Xiao and Ivanova, 2011a) and to develop dose-finding designs that better address the needs of Phase II non-oncology trials (Xiao and Ivanova, 2011b; Ivanova, Xiao and Tymofeyev, 2011).

This dissertation is organized as follows: chapter two provides the literature review; chapter three to five presents the three papers; and chapter six provides a summary of the entire study, study limitations, and conclusions.

CHAPTER 2 LITERATURE REVIEW

This chapter provides a literature review for the three topics: dose-finding when the target dose is on a plateau of a dose-response curve, adaptive isotonic estimation of the minimum effective and peak doses, and two-stage designs for phase II dose-finding trials.

2.1 Dose-finding when the Target Dose is on a Plateau of a Dose-response Curve

Estimating the doses of interest with high precision in proof-of-concept studies is vital for the future development of the drug. The goal of a proof-of-concept study is usually to find the lowest dose with a certain expected target response rate. It is also of interest to compare the response or adverse event rates at the target dose to placebo or active control. Hall et al. (2005) described a proof-of-concept trial for a treatment of migraine headaches. Their study had two goals: first, to find the lowest dose with the response rate of 0.6; second, to compare the response rate with placebo at the estimated dose. Two of the seven scenarios considered by Hall et al. (2005), (0.3,0.3,0.4,0.5,0.6,0.6,0.6) and (0.3,0.6,0.6,0.6,0.6,0.6,0.6), had several doses with the target response rate of 0.6. It is not unlikely in a phase II trial that the response rates will plateau around the rate of interest. When there are several doses with the mean response equal to the target dose, the investigators are usually interested in finding the lowest of these doses. This is because such a dose is likely to have a more favorable adverse event profile.

The second goal in Hall et al. (2005) study was to compare the response rate at the estimated dose with placebo, thus, it is important to maximize the number of subjects at the

estimated target dose. An increasing number of adaptive designs developed for Phase I trials are used in Phase II to target a certain efficacy quantile. As most dose-finding designs were developed and studied for a strictly increasing dose-response curve, it is unknown whether these designs will work well under the possibility of a plateau of a dose-response curve with rates close to the target rate. The continual reassessment method (CRM) can be a good example. The CRM is a dose-finding method proposed by O'Quigley et al. (1990). It has been shown to converge to the target dose (O'Quigley and Shen, 1996) if used in continuous dose space. If used with discrete doses, given a working model, the CRM converges to the target dose or nearby doses with response rates close to the target within the so-called indifference interval (Cheung and Chappell, 2000). Extending the argument from Cheung and Chappell (2000) for a case when a dose-response curve plateaus, the CRM converges to one of the doses on plateau or a dose within indifference interval (Cheung and Chappell, 2000). Therefore, the CRM will yield an increased sample size at one of the doses on the plateau or a nearby dose with response close to the target, although this dose might not be the lowest dose on the plateau.

Let $D = \{d_1,..,d_K\}$ be the ordered set of doses selected for the study. A subject's response at d_k is a Bernoulli random variable with parameter p_k, where $p_1 \le ... \le p_K$. The goal is to find the lowest dose with the response rate Γ. Because we are interested in situations where there is a plateau at the target response rate, we will consider scenarios where $p_1 < p_2 < ... < p_j = ... = p_K = \Gamma$. Note that if the plateau is significantly below or above the target, the designs will perform as well as they do for a strictly increasing curve.

Consider a group design. Subjects are treated in cohorts of size s starting with the lowest dose. Let $X(d_j) \sim \text{Bin}(s, p_j)$ be the number of subjects with response in the most recent

cohort assigned to dose d_j. Let c_L and c_U be two integers such that $0 \le c_L < c_U \le s$. Assume that the most recent cohort of subjects was assigned to dose level d_j, $j = 1, 2\ldots, K$. Then

(1) If $X(d_j) \le c_L$, the next cohort of s subjects is assigned to dose d_{j+1};

(2) If $c_L < X(d_j) < c_U$, the dose is repeated for the next cohort of s subjects;

(3) If $X(d_j) \ge c_U$, the next cohort of s subjects is assigned to dose d_{j-1}.

Appropriate adjustments are made at the lowest and highest doses. The process is continued until N subjects are assigned. This design is denoted as $UD(s, c_L, c_U)$, where s is the cohort size, c_L is the lower cut-off and c_U is the upper cut-off.

Ivanova et al. (2007) described stationary distribution for the group design $UD(s, c_L, c_U)$ when dose-response curve is increasing. We prove a similar result for a non-decreasing curve (Xiao and Ivanova, 2011a).

THEOREM. If the true response rates are $p_1 < p_2 < \ldots < p_j = \ldots = p_K = \Gamma^*$ and the solution of equation (1) for a group design $UD(s, c_L, c_U)$ is equal to Γ^*, the mode of the stationary distribution for the assignments of $UD(s, c_L, c_U)$ spans doses $d_j, \ldots, d_K$.

2.2 Adaptive Isotonic Estimation of the Minimum Effective and Peak Doses

High precision of estimation of doses of interest in dose-ranging studies is essential for evaluating a drug. For example, selecting too high a dose can result in unacceptable toxicity, while choosing too low a dose decreases the chance of showing efficacy in the confirmatory phase, thus reducing the chance of getting regulatory approval of the drug. The minimum effective dose (MED) and the peak dose are the two doses people are mostly interested. The MED is often defined as the lowest dose with response significantly different from placebo, or it can be defined as the dose with the mean response equal to $\mu_0 + \eta$, where

μ_0 is the mean response of placebo and η is the minimum clinically important difference (ICH E4 Guideline, 1994). The peak dose, also sometimes referred to as the maximum useful dose, is the maximum dose beyond which no further beneficial effect is seen (ICH E4 Guideline, 1994). Statistically, we define the peak dose here as the lowest dose with mean response of $\mu_{\max} - \gamma$, where $\mu_{\max}$ is the maximum mean response and γ is a small constant, for example, γ can be set to equal to $0.1(\mu_{\max} - \mu_0)$.

When dose-response curve plateaus near the value of the mean response of interest, the goal is usually to find the lowest dose on the plateau. Cheung (2008) pointed out that existing methods might not be appropriate in this case. For example, the CRM converges to one of those doses but not necessarily the lowest one. The stationary distribution of a group design (Wetherill, 1963) will be uniformly spread across all target doses (Xiao and Ivanova, 2011a). That is, existing adaptive methods will not work for estimating the peak dose, the lowest dose on the plateau. Also, existing methods have not been designed for the case when the target is defined based on the mean response at one of the doses, for example, at placebo.

Isotonic assumptions can be made in most dose-finding trials. Utilizing the isotonic assumption usually leads to increased efficiency in the estimation of the target dose compared to a trial where this assumption is not utilized. Our investigation shows that this is especially true if the dose-response curve has a plateau. Researchers have successfully estimated the dose-response curve under an isotonic model without assuming a parsimonious model for dose-response relationship in several recent publications. For example, Conaway et al. (2004), Yuan and Chappell (2004), Ivanova and Wang (2006) and Ivanova and Kim (2009) proposed frequentist methods; Li et al. (2008) and Bekele et al. (2008) proposed Bayesian approaches for various dose-finding problems.

Often a set of known covariates are believed to be associated with response to treatment. For an example, a large pharmaceutical company recently conducted a proof-of-concept study in which the in- and out-patient status was believed to be associated with therapeutic response to treatment. Researchers have proposed a number of adaptive methods to address the problem of estimating the target dose for each level of covariate (O'Quigley and Paoletti, 2003; Ivanova and Wang, 2006). A more challenging problem is to find dose for each subject according to his/her covariate values when the outcome is binary. A recent publication by Thall et al. (2009) addressed this problem. When a target dose is defined using a reference dose (for example, placebo) and when the mean response is modeled with identity link function using a linear model with covariates, the target doses for different levels of covariate coincide. In such trials the role of covariates is similar to that in a comparative multi-arm trial: balancing with respect to covariates is preferred (Atkinson, 1999) for validity and increased efficiency of estimation. Balancing is more challenging in the context of an adaptive dose-finding trial compared to a parallel group study.

We will investigate the optimal allocation when the MED and peak doses are estimated using isotonic regression; we then use the knowledge of optimal allocation to construct sequential dose-finding design to estimate the MED and peak doses (Xiao and Ivanova, 2011b). We will also describe how to randomize subjects to doses in the course of adaptive trial while balancing allocations with respect to known covariates (Xiao and Ivanova, 2011b).

2.3 Dose-finding when the Target Dose is on a Plateau of a Dose-response Curve

Efficacy is usually the most important thing to consider in selecting candidate doses of a drug in Phase II dose-finding trials, and estimating MED is almost always one of the

goals of the trial. Often times both efficacy and safety are taken into consideration when selecting the best dose, as increasing the dose results in both higher efficacy and increased toxicity or adverse event rates. The common approach is to quantify the trade-off between efficacy and the adverse event rate through a utility function. Such function incorporates both efficacy and safety into a measure of overall clinical "utility" (Thall et al., 2002; Ivanova, 2003; Dragalin and Fedorov, 2006; Thall and Cook, 2004; Berry et al., 2001; Fedorov and Wu, 2007; Ivanova et al., 2009). That utility function often has an "umbrella" or "inverse U" shape. The objective of a trial, therefore, is to maximize the overall clinical "utility" of the drug. We will refer to the dose that maximizes the utility function as the optimal dose.

An additional objective may be to test efficacy and adverse event rates at the MED or the optimal dose against placebo and/or an active control. Therefore, a good assignment strategy for a dose-ranging study will be the one that provides both a good-quality estimation of the target dose and the increased sample size at the estimated target dose to yield better power of treatment comparisons.

Most dose-finding designs (one notable exception is the design in Berry et al., 2001) assume a certain order of the means of the dose-response curve, such as non-decreasing, umbrella, or various partial orders. Isotonic estimates have been successfully used in adaptive dose-finding studies by Conaway et al. (2004), Yuan and Chappell (2004), and more recently by Li et al. (2008) and Bekele et al. (2008) in the context of a Bayesian dose-finding trial.

Most dose-finding designs are fully sequential designs proposed for oncology studies where the experimentation starts at the lowest dose and the dose is gradually escalated because of ethical considerations. That is probably why most of the adaptive dose-finding

methods for non-oncology dose-finding studies are also fully sequential or multi-stage designs (Berry et al., 2001; Ivanova et al., 2008; Ivanova et al., 2009).

However, the logistics of implementing a fully sequential dose-finding study can be daunting. This is one reason why the most common design in dose-ranging studies is equal allocation to all doses. A two-stage design seems to be a reasonable compromise between a multi-stage and a single-stage approach. Miller et al. (2007) recently investigated a two-stage strategy for a dose-ranging study that is optimal across several parametric models. They concluded that the proposed two-stage strategy offers minor benefit compared to a single-stage design in terms of the efficiency of estimating the target dose. Dragalin and Fedorov (2006) investigated the optimal two-stage designs for two correlated binary endpoints that follow a bivariate probit model and concluded that the two-state strategy is superior to equal allocation.

We propose Bayesian two-stage designs for dose-ranging trials under the following three models: estimating MED under assumption of non-decreasing dose-response curve, estimating the dose with the highest response under umbrella order assumption, estimating MED under isotonic matrix order (Ivanova et al., 2011). The latter problem arises when several different administration schedules are investigated.

CHAPTER 3 DOSE-FINDING WHEN THE TARGET DOSE IS ON A PLATEAU OF A DOSE-RESPONSE CURVE

Consider a problem of estimating a dose with certain response rate. Most dose-finding designs for this problem were developed and studied in cases where the mean dose-response is strictly increasing in dose. In Phase II dose-finding studies often the dose-response curve plateaus in the range of interest and there are several doses with the mean response equal to target. In this case it is usually of interest to find the lowest of these doses since higher doses might have higher adverse event rates. It is often desirable to compare the response rate at the estimated target dose with placebo and/or active control. We investigate which of the several known dose-finding methods is the most suitable when dose response curve plateaus. Some of the designs tend to spread the allocation among the doses on the plateau, others, like the continual reassessment method and the *t*-statistic design, concentrate allocation at one of the doses with the *t*-statistic design selecting the lowest dose on the plateau more frequently.

KEY WORDS: Proof-of-concept; Phase II trials; Group up-and-down designs; Continual reassessment method; *t*-statistic design.

3.1 Introduction

Estimating the doses of interest with high precision in proof-of-concept studies is vital for the future development of a drug. The goal of a proof-of-concept study is often to find the lowest dose with a certain expected target response rate. Often it is also of interest to

compare the response or adverse event rates at the target dose to placebo or active control. Hall, Meier, and Diener (2005) described a proof-of-concept trial for a treatment of migraine headaches. There were two goals in the study. The first goal was to find the lowest dose with the response rate of 0.6, and the second goal was to compare the response rate at the estimated dose with placebo. Two of the seven scenarios considered by Hall et al. (2005), (0.3,0.3,0.4,0.5,0.6,0.6,0.6) and (0.3,0.6,0.6,0.6,0.6,0.6,0.6), had several doses with the target response rate of 0.6. It is not unlikely in a Phase II trial that the response rates will plateau around the rate of interest. When there are several doses with the mean response equal to target, the investigators are interested in finding the lowest of these doses. One of the reasons is because such a dose is likely to have a more favorable adverse event profile. To achieve the second goal in Hall et al. (2005) study, it is important to maximize the number of subjects at the estimated target dose. More and more adaptive designs developed for Phase I trials are used in Phase II to target a certain efficacy quantile. As most of dose-finding designs were developed and studied for a strictly increasing dose-response curve, it is not known which of these designs will work the best when there is a possibility of a plateau of a dose-response curve with rates close to the target rate. In this paper we investigate, theoretically and via simulations, the performance of several dose-finding designs such cases. The designs studied are group designs (Wetherill, 1963; Ivanova, 2006), the continual reassessment method (CRM) (O'Quigley, Pepe , and Fisher, 1990), and the dose-finding design based on *t*-statistic (Ivanova and Kim, 2009).

3.2 Group Designs

Let $D = \{d_1,..,d_K\}$ be the ordered set of doses selected for the study. A subject's response at d_k is a Bernoulli random variable with parameter p_k, where $p_1 \le ... \le p_K$. The

goal is to find the lowest dose with the response rate Γ. Because we are interested in situation where there is a plateau at the target response rate, we will consider scenarios where $p_1 < p_2 < ... < p_j = ... = p_K = \Gamma$. Note that if the plateau is significantly below or above the target, the designs will perform as well as they perform for a strictly increasing curve.

First we consider a group design, as this was selected by the investigators of the migraine headache trial (Hall et al., 2005). Subjects are treated in cohorts of size s starting with the lowest dose. Let $X(d_j) \sim \text{Bin}(s, p_j)$ be the number of subjects with response in the most recent cohort assigned to dose d_j. Let c_L and c_U be two integers such that $0 \le c_L < c_U \le s$. Assume that the most recent cohort of subjects was assigned to dose level $d_j, j = 1, 2\ldots, K$. Then

(1) If $X(d_j) \le c_L$, the next cohort of s subjects is assigned to dose d_{j+1};

(2) If $c_L < X(d_j) < c_U$, the dose is repeated for the next cohort of s subjects;

(3) If $X(d_j) \ge c_U$, the next cohort of s subjects is assigned to dose d_{j-1}.

Appropriate adjustments are made at the lowest and highest doses. The process is continued until N subjects are assigned. This design is denoted as $UD(s, c_L, c_U)$, where s is the cohort size, c_L is the lower cut-off and c_U is the upper cut-off.

Hall et al. (2005) used $UD(4,2,3)$ in the headache trial, where the dose is increased if 2 or less responses are observed, and the dose is reduced if 3 or more responses were observed (Hall et al., 2005). When the dose-response curve is strictly increasing, $p_1 < ... < p_K$, the assignments in a group design will cluster around the dose with response rate Γ^* (Ivanova, Flournoy and Chung, 2007), where Γ^* is the solution of

$$\Pr\{\text{Bin}(s, \Gamma^*) \le c_L\} = \Pr\{\text{Bin}(s, \Gamma^*) \ge c_U\}. \quad (1)$$

For *UD*(4,2,3), $\Gamma^* = 0.6143$. Therefore, it was appropriate to use *UD*(4,2,3) to target the response rate of $\Gamma = 0.6$. See Ivanova et al, 2007 for guidelines on how to choose design parameters s, c_L, c_U to target desired quintile Γ. The investigators of the trial justified using *UD*(4,2,3) for $\Gamma = 0.6$ via simulations. The question is how well this group design behaves if the condition $p_1 < ... < p_K$ is violated. The theorem below states that for a dose-response curve with $p_1 < p_2 < ... < p_j = ... = p_K = \Gamma^*$, for large sample size N the assignments will be equally spread over doses $d_j, ..., d_K$ rather than concentrating on one of these doses.

THEOREM. If the true response rates are $p_1 < p_2 < ... < p_j = ... = p_K = \Gamma^*$ and the solution of equation (1) for a group design *UD*(s, c_L, c_U) is equal to Γ^*, the mode of the stationary distribution for the assignments of *UD*(s, c_L, c_U) spans doses $d_j, ..., d_K$. The proof of the theorem is in the Appendix A.

For example, if response rates at the doses are (0.3, $\Gamma^*, \Gamma^*, \Gamma^*, \Gamma^*, \Gamma^*, \Gamma^*$) with $\Gamma^* = 0.6143$, and the total sample size in the trial is relatively large, the proportions of subjects allocated to the seven doses in the limit by *UD*(4,2,3) is $(\pi_1, \pi, \pi, \pi, \pi, \pi, \pi)$ with $\pi_1 = 0.082$ and $\pi = 0.153$. The more doses are on the plateau, the smaller the proportion of subjects allocated to each of the doses on the plateau and the smaller the power of comparison with placebo.

3.3 Dose Finding Based on *t*-statistic

The *t*-statistic design was proposed by Ivanova and Kim (2009). It can be used with any type of outcome. Let $\boldsymbol{n}(t) = (n_1(t), ..., n_K(t))$ be the number of subjects at each of the K doses right after subject t, $t \le N$, has been assigned, that is, $n_1(t) + ... + n_K(t) = t$. Let Y_{ji} be

the observation from the *i*th subject assigned to dose d_j, $i = 1,..,\ n_j(t)$. Let $\hat{p}_j = \sum_{i=1}^{n_j(t)} Y_{ji} / n_j(t)$ be the current estimate of response rate at dose d_j, computed from all subjects assigned to d_j so far. Define $T_j\left(n_j(t)\right),\ n_j(t) = 2,3...$, to be the *t*-statistic

$$T_j\left(n_j(t)\right) = \frac{\hat{p}_j - \Gamma}{\sqrt{p_j(1-p_j)} / \sqrt{n_j(t)}}.$$

If $\hat{p}_j = 0$ or 1, $T_j\left(n_j(t)\right)$ is equal to $+\infty$ or $-\infty$ depending on the sign of $\hat{p}_j - \Gamma$. Subjects are assigned in cohorts or one at a time. Suppose the most recent subject *t* was assigned to dose d_j. The next subject is assigned as follows:

(i) if $T_j\left(n_j(t)\right) \le -\Delta$, the next subject is assigned to dose d_{j+1};

(ii) if $T_j\left(n_j(t)\right) \ge \Delta$, the next subject is assigned to dose d_{j-1};

(iii) if $-\Delta < T_j\left(n_j(t)\right) < \Delta$, the next subject is assigned to dose d_j.

Ivanova and Kim (2009) recommended to set design parameter $\Delta = 1$. The performance of the *t*-statistic design where there is a plateau in the range of interest is assessed by simulations in Section 3.5.

3.4 The Continual Reassessment Method (CRM)

The CRM is a dose-finding method proposed by O'Quigley et al. (1990). It uses a working model for dose-response relationship, for example, where $p_i = b_i^{\theta}$, where $(b_1,...,b_K)$ is a set of constants and θ is a parameter to be estimated. The CRM has been shown to converge to the target dose (O'Quigley and Shen, 1996) if used in continuous dose space. If used with discrete doses, given a working model, the CRM converges to the target dose or nearby doses with response rates close to the target within so called indifference interval.

Extending the argument from Cheung and Chappell (2000) for a case when a dose-response curve plateaus, the CRM converges to one of the doses on plateau or a dose within indifference interval. Therefore, the CRM will yield an increased sample size at one of the doses on the plateau or a nearby dose with response close to Γ, although this dose might not be the lowest dose on the plateau. In the simulation study we used $(b_1,...,b_7)$ = (0.1,0.2,0.3,0.4,0.5,0.6,0.7) and exponential prior with mean 1 for parameter θ.

3.5 Simulation Study

In addition to the two scenarios from Hall et al. (2005), four more scenarios were used to compare the different designs in our simulation study (Table 1). The target treatment response rate is Γ = 0.6, and placebo response rate is 0.3.

We investigated the performance of the group design *UD*(4,2,3) used in Hall et al. (2005), the CRM and the *t*-statistic design. The sample size was fixed at 120 subjects with 40 subjects assigned to placebo and 80 to various doses of the drug. This sample size was chosen because 40 subjects per group yield 80% power if treatments with true rates of 0.3 and 0.6 are compared using one-sided 0.05 level test. Subjects were assigned in cohorts of 6 with 2 subjects assigned to placebo and 4 to a dose of the drug. At the end of the trial, response rates in the *t*-statistic and the group designs were estimated using isotonic regression (Barlow, Bartholomew, Bremner, and Brunk, 1972) and then the dose with the estimated response rate closest to the target was declared the estimated target dose. If there were two or more such doses, the highest dose with the estimated value below Γ was chosen. If all the estimated values at these doses were higher than Γ, the lowest of these doses was chosen. For the CRM design, the estimated target dose was defined as in O'Quigley et al. (1990) as the dose that would have been recommended for the next subject.

Results for each design/scenario combination are based on 5000 simulation runs. The primary goal of our study was to estimate the target dose. Table 2 showed the selection probability of each dose as the estimated target dose under each design. All designs perform well in selecting a dose with the target response rate, not necessarily the lowest, with the CRM performing the best. The results for the lowest dose on the plateau are in bold. The *t*-statistic design selects the lowest dose on the plateau more frequently than others in scenarios 1 – 4, and the CRM performs the best in Scenario 6 where the dose-response curve is strictly increasing. The average sample size at each dose is shown in Table 3, with the results for the lowest dose on the plateau shown in bold.

Another goal of the trial was to compare the response rate at the estimated target dose with the placebo. That is why it is important to have a large sample size at the estimated target dose. The distributions of the sample size at the estimated target dose are displayed in Table 4. The CRM and the *t*-statistic design have larger average sample size at the estimated target dose compared to the group design.

The lowest dose on the plateau is often of interest because it is likely to have better adverse event profile than higher doses. We constructed plausible adverse event rate scenarios (Table 1), and compared the estimated dose with placebo based on both efficacy and adverse event rates. The adverse event rate at the estimated dose was compared with the rate of placebo using one sided test with the null hypothesis that the adverse event rate of the drug is higher than placebo rate plus 0.2. Table 5 displays the proportion of trials where the estimated target dose is shown to have the efficacy rate significantly better than placebo rate and adverse event rate significantly lower than placebo rate plus 0.2. As far as power for joint comparison, the *t*-statistic design performs the best because it yields large sample size at the

estimated target dose on average and also selects the lowest dose on the plateau more often than other designs.

3.6 Conclusion

It is not uncommon for a dose-response curve to plateau, yielding several doses with the same mean response. We are not aware of existing designs that target the lowest dose on the plateau of a dose response curve. We investigated the performance of several known dose-finding designs developed for strictly increasing curves in the case when a dose-response curve plateaus. Both the CRM and the *t*-statistic design performed well with the *t*-statistic design selecting the lowest dose on the plateau more frequently. We have demonstrated theoretically and by simulations that a group design is not a good design choice if it is likely that a dose-response curve plateaus near the response rate of interest. Until designs to target the lowest dose on the plateau are developed, we recommend using the CRM if the goal is to find any dose on the plateau with a certain mean response. We recommend using the *t*-statistic design if it is of interest to find the lowest dose on the plateau.

We considered the case where a dose-response curve is assumed to be non-decreasing. The methods we have studied are not appropriate when there might be a downturn of a dose-response curve at higher doses such is in Ivanova, Liu, Snyder and Snavely (2009), where a dose-finding method that works with umbrella shaped efficacy curve and incorporates toxicity is proposed.

CHAPTER 4 ADAPTIVE ISOTONIC ESTIMATION OF THE MINIMUM EFFECTIVE AND PEAK DOSES

We obtain the optimal allocation for estimating the minimum effective and peak doses in a dose-ranging trial when the set of dose level is fixed and isotonic regression is used as a method of estimation. We propose fully sequential strategy for subject assignment. The proposed strategy includes adaptive randomization procedure to balance the allocation to placebo and active doses. We also consider estimation in presence of covariates, in which case randomization procedure also balances the allocation with respect to a set of known covariates.

KEY WORDS: Dose-ranging; Minimum effective dose; Peak dose; Phase II trials; Up-and-down designs.

4.1 Introduction

High precision of estimation of doses of interest in dose-ranging studies is essential for evaluating a drug. The minimum effective dose (MED) and the peak dose are the two doses most of interest. The MED is the smallest dose with a discernible useful effect (ICH E4 Guideline, 1994). The MED is often defined as the lowest dose with response significantly different (referring to statistical significance) from placebo. Alternatively it can be defined in continuous dose space as the dose with the mean response equal to $\mu_0 + \eta$, where μ_0 is the mean response of placebo and η is the minimum clinically important difference. The MED may not exist, if mean response at all doses in the range studied is less than $\mu_0 + \eta$. The peak

dose, also sometimes referred to as the maximum useful dose, is a maximum dose beyond which no further beneficial effect is seen (ICH E4 Guideline, 1994). Locating the peak dose is usually of interest after the drug was shown to be efficacious. The peak dose is the lowest dose on the plateau of a dose-response curve. Mathematically, in continuous dose space we define the peak dose here as the lowest dose with mean response of $\mu_{\max} - \gamma$, where $\mu_{\max}$ is the maximum mean response and γ is a small constant.

There is a long history of adaptive dose-finding methods for estimating a dose with a certain mean response when outcome is binary (e.g. Wetherill, 1963; O'Quigley et al., , 1990; Babb et al., 1998) and for continuous outcomes (e.g. Eichhorn and Zacks, 1973; Ivanova and Kim, 2009). All of these methods have been developed under the assumption that the mean response is strictly increasing with dose. When dose-response curve plateaus near the value of the mean response of interest, the goal is usually to find the lowest dose on the plateau. Cheung (2008) pointed out that existing methods might not be appropriate in this case, for example, the CRM converges to one of such doses and not necessarily the lowest one (Xiao and Ivanova, 2011). The stationary distribution of a group design (Wetherill, 1963) will be uniformly spread across all target doses (Xiao and Ivanova, 2011). That is, existing adaptive methods will not work for estimating the peak dose, the lowest dose on the plateau. Also, existing methods have not been designed for the case when finding the target dose requires estimating mean responses at other doses, for example, finding the location of the MED requires estimation of placebo response.

We make an assumption that the mean response is non-decreasing with dose. Such isotonic assumptions can be made in most of dose-finding trials. Using the isotonic assumption usually leads to increased efficiency in the estimation of the target dose

compared to a trial where this assumption is not utilized. Our investigation shows that this is especially true if the dose-response curve has a plateau. Isotonic estimates were successfully used in adaptive dose finding by Conaway et al. (2004); Yuan and Chappell (2004); and recently by Li et al. (2008) and Bekele et al. (2008) in the context of a Bayesian dose-finding trial.

Often there is a set of known covariates that are believed to be associated with response to treatment. Our motivating example is a recent Phase II dose-finding trial conducted by a large pharmaceutical company where it was believed that in- and out-patient status were associated with therapeutic response to treatment. A number of adaptive methods address the problem of estimating the target dose for each level of covariate (e.g., O'Quigley and Paoletti, 2003; Ivanova and Wang, 2006). A recent publication by Thall et al. (2009) addressed a rather challenging problem of dosing each subject according to his/her covariate values when the outcome is binary. Both the MED and the peak dose are defined using a reference dose, placebo or the highest dose. When defined this way, the target dose will not depend on covariates as long as effects of dose and covariates are additive (no interaction). In dose-finding trials the role of covariates is similar to that in a comparative multi-arm trial: balancing with respect to covariates is preferred (Atkinson, 1999) for validity and to increase efficiency of estimation. Balancing is more challenging in the context of an adaptive dose-finding trial compared to a parallel group study. We describe how to randomize subjects to doses in the course of adaptive trial while balancing allocations with respect to known covariates.

4.2 Optimal Allocation for Estimating the MED and Peak Doses

4.2.1 Notation.

Let $\{d_0,...,d_K\}$ be the set of ordered dose levels selected for a trial with d_0 denoting placebo and d_K denoting the highest dose, for example, the maximum tolerated dose established in earlier trials. Let n be the total sample size, let n_i be the number of subjects assigned to d_i at the time a total of n subjects are assigned, $n_0+...+n_K=n$. Let Y_{ij} denote the response of the jth subject, j = 1, 2,..., n_i, assigned to d_i, i = 0, 1,..., K. Let x_{ij} be a $K\times 1$ vector of covariates associated with the jth subject assigned to the ith dose. Consider a linear model

$$Y_{ij}=\mu_i+x_{ij}'\beta+\varepsilon_{ij}, i=0,1,...,K, j=1,2,...,n_i. \qquad (1)$$

Here μ_i is the mean response at d_i, β is the regression parameter associated with covariate vector x_{ij} and $\varepsilon_{ij}\sim N(0,\sigma^2)$.

The MED is defined as the dose with the mean response of $\mu_0+\eta$, where $\eta>0$ is the minimum clinically important difference specified before the trial. The peak dose is defined as the lowest dose with the mean response of $\mu_K-\gamma$, where γ, $\gamma\geq 0$, reflects the proximity to the highest mean response. For example, in a seven-dose trial with sigmoid dose-response curve with true mean responses at seven doses of (0.2,0.21,0.25,0.5,0.74,0.79,0.8), the MED defined with $\eta=0.35$ is d_3 and the peak dose defined with γ = 0.06 is dose d_4. If the dose-response curve is flat, the peak dose will coincide will placebo. It is of interest to locate the peak dose only if there is dose response and therefore the peak dose is as high or higher than the MED.

4.2.2 Isotonic estimation of the target dose.

Let $\hat{\boldsymbol{\mu}} = \left(\hat{\mu}_0^U, ..., \hat{\mu}_K^U\right)$ be the vector of maximum likelihood estimates (MLE) obtained from model (1) where *U* stands for unconstrained estimator, and Σ be its variance covariance matrix. The constrained MLE, $\hat{\boldsymbol{\mu}} = \left(\hat{\mu}_0, ..., \hat{\mu}_K\right)$, is the estimator that maximizes the likelihood according to model (1) under the restriction $\hat{\mu}_0 \le ... \le \hat{\mu}_K$. When $\left(\hat{\mu}_0^U, ..., \hat{\mu}_K^U\right)$ are independent, the constrained MLEs can be computed by applying the pool adjacent violator algorithm to unconstrained estimates (Robertson et al., 1988). That is, if $\hat{\mu}_0^U \le ... \le \hat{\mu}_K^U$, $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}^U$; otherwise, the data from adjacent doses where the assumption of monotonicity is violated are pooled (see Robertson et al., 1988, or Stylianou and Flournoy, 2002, for more details). In presence of covariates the pool adjacent violator algorithm applied to unconstrained MLEs will not yield the constrained MLEs and might result in the estimates with increased mean squared error (Hwang and Peddada, 1994). A projection approach that takes covariance into account (Silvapulle and Sen, 2005) cannot be used here since the variance is not known. We computed constrained MLEs directly by maximizing the likelihood in (1) (see section 7.2.1, p 246 of McCullagh and Nelder, 1989, for details).

Further, we define two estimators based on $\hat{\boldsymbol{\mu}}$. The first will be referred to as the *lowest dose* estimator and is defined as the lowest dose on the plateau of doses with the estimated mean response closest to target. This estimator is suitable for estimating the peak dose. For example, if $\hat{\boldsymbol{\mu}}$ = (0.22,0.22,0.45,0.45,0.76,0.76,0.76), the estimated mean response closest to $\hat{\mu}_6 - \gamma = \mu_6 - 0.06 = 0.7$ is $\hat{\mu}_4 = \hat{\mu}_5 = \hat{\mu}_6 = 0.76$, and the lowest dose estimator will select d_5 as the estimated peak dose. The second estimator is referred to as the *closest dose*

estimator and is suitable for estimating the MED. It selects the lowest dose among doses with the mean response closest to target if their estimated mean is higher than the target; and selects the highest of the doses if their estimated mean is lower than the target. In the example above, when estimating the MED with $\eta = 0.3$ the estimated mean response closest to $\hat{\mu}_0 + 0.3 = 0.52$ is $\mu_2 = \mu_3 = 0.45$, and the closest dose estimator will select the highest dose on the plateau, d_3, as the estimated MED since 0.45 < 0.52.

4.2.3 Optimal allocation to estimate the MED and Peak Dose.

When developing an adaptive allocation it is important to know which fixed allocation is the most efficient for estimating the target dose. An optimal design is an allocation that optimizes a certain criterion with respect to the proportion of subjects $(w_0,...,w_K)$, $w_i \geq 0$, assigned to each support point, dose, $(d_0,...,d_K)$. The classical optimal design (Pukelsheim, 1993) in continuous dose space optimizes a certain quantity of interest, such as the volume of the confidence ellipsoid (D-optimal design) or the average variance of parameter estimates (A-optimal design). With discrete dose space, it is most natural to maximize the probability of correct selection of the target dose. Since in most of dose-finding trials we work with a set of doses that have been selected before the trial, we are concerned with identifying the optimal weights for the doses. In most cases, the optimal design depends on true model parameters that are not known before the experiment. This is true in case of isotonic estimation as well.

In the result below $s = 0$ and $\nu = \eta$ when the MED is being estimated; $s = K$ and $\nu = -\gamma$ when the peak dose is being estimated. The following is true (the proof is in Appendix B).

Proposition. *The probability of correctly selecting the target dose defined as the dose with the mean response equal to* $\mu_s + \nu$, $s \in \{0, K\}$ *by applying the closest dose or the lowest*

dose estimator to the weighted average of components of $\hat{\boldsymbol{\mu}}^U$ *depends only on* $\{w_i\}$, $\left\{(\mu_i - \mu_s)\sqrt{n}/\sigma\right\}$ *and* $\left(n/\sigma^2\right)\Sigma$.

When the pooled adjacent violator algorithm is used resulting isotonic estimates are the weighted average of components of $\hat{\boldsymbol{\mu}}^U$, hence it follows from the Proposition that the optimal weights are a function of $\left\{(\mu_i - \mu_s)\sqrt{n}/\sigma\right\}$ and $\left(n/\sigma^2\right)\Sigma$. In case when $\hat{\boldsymbol{\mu}}^U$ are not correlated, the optimal weights are a function of $\left\{(\mu_i - \mu_s)\sqrt{n}/\sigma\right\}$ only.

Consider the problem of estimating a dose with a mean response ν, where ν is a known constant, and when no covariates are present. We use a two-step approach to compute the optimal design. In the first step we determine which support points have non-zero weight. Then, compute optimal weights for these support points. In the first step, for given $\left\{\mu_i\sqrt{n}/\sigma\right\}$ we compute the optimal design numerically using the Nelder-Mead algorithm (Nelder and Mead, 1965). In all dose-response scenarios the optimal design for the lowest dose estimator is at most three-point design with allocations to $d_{\tau-1}$, d_τ, and $d_{\tau+1}$, where d_τ is the true target dose. Moreover, unless $\mu_{\tau+1} - \mu_\tau$ is very large, the optimal design for the lowest dose estimator is a two-point design with allocations to $d_{\tau-1}$ and d_τ. Since in the peak dose estimation a dose-response curve plateaus and $\mu_{\tau+1} - \mu_\tau \le \gamma$, where γ is small, the optimal design to estimate the peak dose using the lowest dose estimator is a three-point design with allocation to $d_{\tau-1}$, d_τ and d_K, where d_τ is the true peak dose. For the closest dose estimator the optimal design is at most a three-point design with non-zero weights at the true target dose, d_τ, the dose right below, $d_{\tau-1}$, and the dose right above, $d_{\tau+1}$. Therefore the

optimal design to estimate the MED using the closest dose estimator is at most a four-point design with allocation to d_0, $d_{\tau-1}$, d_τ, and $d_{\tau+1}$, where d_τ is the true MED.

In the second step of the optimal design calculations, we use normal cumulative distribution function to compute optimal weights. The optimal weights to estimate the peak dose are computed based on $\left\{(\mu_\tau - \mu_K)\sqrt{n}/\sigma\right\}$ and $\left\{(\mu_{\tau-1} - \mu_K)\sqrt{n}/\sigma\right\}$. The probability of correctly selecting d_i as the estimated target dose is equal to

$$P = \Pr\left\{\hat{\mu}_{i-1} < \mu_\tau^U < \hat{\mu}_K \quad \cap \quad \mu_{\tau-1}^U + \hat{\mu}_i < 2\left(\mu_K^U - \gamma\right) < \hat{\mu}_i + \mu_K^U\right\}$$
$$+ \Pr\left\{\hat{\mu}_K < \mu_\tau^U \quad \cap \quad \hat{\mu}_{i-1} + \mu_{\tau,K}^U < 2\left(\hat{\mu}_{\tau,K}^U - \gamma\right)\right\},$$

where $\hat{\mu}_{\tau,K}^U$ denotes the sample mean of a pooled sample obtained at d_τ and d_K.

Equivalently,

$$P = \Pr\left\{\left(\hat{\mu}_{i-1} - \mu_K^U\right) - \left(\hat{\mu}_i - \mu_K^U\right) < 0 < \hat{\mu}_K - \mu_\tau^U < 2\gamma < \left(\hat{\mu}_K - \mu_\tau^U\right) + \left(\hat{\mu}_K - \mu_{\tau-1}^U\right)\right\}$$
$$+ \Pr\left\{0 < \hat{\mu}_i - \mu_K^U \quad \cap \quad 2\gamma < \hat{\mu}_i \frac{w_\tau}{w_\tau + w_K} + \mu_K^U \frac{w_K}{w_\tau + w_K} - \hat{\mu}_{\tau-1}^U\right\}. \tag{2}$$

Note that expression to the right of 2γ can be written as a function of $(\hat{\mu}_{i-1} - \mu_K^U, \hat{\mu}_i - \mu_K^U)$. Vector $(\hat{\mu}_{i-1} - \mu_K^U, \hat{\mu}_i - \mu_K^U)$ follows bivariate normal distribution with mean vector $(\mu_{\tau-1} - \mu_K, \mu_\tau - \mu_K)$ and variance covariance matrix with diagonal $\sigma^2/n\left(1/w_{\tau-1} + 1/w_K, 1/w_\tau + 1/w_K\right)$ and off-diagonal elements $-\sigma^2/\left(w_K n\right)$. The probability *P* can be computed using cumulative function of the multivariate normal distribution. The optimal allocation $(w_{\tau-1}, w_\tau, w_K)$ is the one that maximizes *P* over $(w_{\tau-1}, w_\tau, w_K)$, $0 \le w_k \le 1$, $k = \tau - 1, \tau, K$. We computed the optimal allocation using the Nelder-Mead algorithm. Figure 1 displays $(w_{\tau-1}, w_\tau, w_K)$ plotted against total sample size for $(\mu_{\tau-1}, \mu_\tau, \mu_K) =$

$(0.5, 0.74, 0.8)$ with $\sigma = 0.25$. This mean vector is one of the scenarios in Bretz et al. (2005) (Table 6, scenario 5). The optimal allocation for sample sizes larger than in a typical trial is displayed to illustrate that, for the set-up considered, the optimal allocation proportion to $d_{\tau-1}$ gets smaller and the allocations to d_τ and d_K increase as $\sigma^2/\sqrt{n}$ gets smaller. This is because μ_τ is closer to μ_K than it is to $\mu_{\tau-1}$, therefore it is more efficient to spend resources on distinguishing between μ_τ and μ_K, than between μ_τ and $\mu_{\tau-1}$. The probability of correctly identifying the target dose in the range of sample sizes of interest for optimal allocation is similar to equal allocation to the three doses with minimum relative efficiency of

$$\min_{n\in[20,100]} P(1/3,1/3,1/3)/P(w_{\tau-1}^{opt}, w_\tau^{opt}, w_K^{opt}) = 0.99 .$$

That is, to estimate the peak dose well, we need to assign about equal number of subjects to the peak dose, a dose right below and the highest dose with no assignments to other doses. Interestingly, unbalanced allocation with many more subjects assigned to the peak dose is only beneficial when the standard error of the estimated mean is very small compared to the difference between means. Also, in the estimation of the peak dose increased allocation to the doses on the plateau other than the peak and the high dose, substantially decrease the precision of the estimation of the peak dose.

The optimal weights for the four-point design to estimate the MED are calculated similarly. Figure 2 displays $(w_0, w_{\tau-1}, w_\tau, w_{\tau+1})$ plotted against total sample size for $(\mu_0, \mu_{\tau-1}, \mu_\tau, \mu_{\tau+1}) = (0.2, 0.25, 0.5, 0.74)$ with $\sigma = 0.25$ (Table 6, scenario 5). The conclusion is similar, allocating approximately equal number of subjects to each of the four doses yields good quality of estimation of the target dose. As $\sigma^2/\sqrt{n}$ gets smaller, the allocation proportions to $d_{\tau-1}$ and $d_{\tau+1}$ decrease and the allocations to d_0 and d_τ increase.

Optimal allocations in Section 4.2 were computed under the assumptions of independence among unconstrained estimates. When covariates are present the unconstrained estimates are no longer independent, and optimal allocations are computed under given correlation structure. Our simulation study of trials with covariates yielded similar optimal designs: three or four point designs with balanced allocation are nearly optimal.

4.3 Adaptive Design to Estimate the MED and Peak Doses

4.3.1 Adaptive strategy to estimate the MED and the Peak Dose.

In Section 4.2 we computed the optimal design for estimating the MED and peak doses. As is the case for most parametric models, the optimal allocation depends on the true model. For isotonic model, one needs to know the location of the target dose to construct the optimal design. Therefore our adaptive strategy will be to locate the target dose and to have allocations to the target dose and other two or three key doses approximately equal. We use this as a guideline to design an adaptive strategy. Ivanova and Kim (2009) introduced a dose-finding design based on *t*-statistic to locate the dose with a certain mean response. We modify their strategy to target optimal (or nearly optimal) allocation. According to the *t*-statistic design, subjects can be assigned in groups or one at a time. Assume that the most recent assignment was to dose d_i. Let T_i be the test statistic testing H_0: $\mu_i - (\mu_0 - \eta) = 0$ against the two-sided alternative computed using constrained MLEs $\hat{\mu}_i$, $\hat{\mu}_K$ and the estimated common variance from linear model (1) . Then,

(i) If $T_i \le -\Delta$, the next group of subjects is assigned to doses d_{i+1};

(ii) If $-\Delta < T_i < \Delta$, the next group of subjects is assigned to doses d_i;

(iii) If $T_i \ge \Delta$, the next group of subjects is assigned to doses d_{i-1}.

Applying this rule when the current dose is d_1 or d_K might cause the dose assignment to be outside $\{d_1,...,d_K\}$. Thus for i = 1 or K, when the rule would cause a treatment to be outside of the dose levels, the current dose is repeated instead.

To estimate the MED, we set $\Delta = 0$, in the above design in which case the adaptive rule will allocate to either d_{i-1} or d_{i+1} after allocating d_i. If $\mu_\tau = \mu_0 - \eta$, the limiting allocation for the *t*-statistic design is allocating to $d_{\tau-1}$, d_τ, and $d_{\tau+1}$ with proportion (0.25,0.5,0.25). This proposed strategy provides acceptable balance in allocations to $d_{\tau-1}$, d_τ, and also allows the design to "move fast" among doses in the early stages of the trial (Ivanova and Kim, 2009).

To estimate the peak dose, one needs to make sure that the design converges to the lowest dose on the plateau and also that the allocation close to optimal is achieved. We accomplish this by a choice of Δ and by modified the decision rule in the design. To make sure that the design reaches the lowest dose of the plateau, we replace the action "if $-\Delta < T_i < \Delta$, repeated the dose" in the design described above with the action "if $-\Delta < T_i < \Delta$, assign next subject to d_{i-1} with some probability φ or repeat the dose with probability 1-φ while keeping Δ strictly above 0. Ivanova and Kim (2009) pointed out that it is advantageous to have small Δ in the beginning of the trial for "fast movement" with larger Δ later in the trial. For example, a trial with 8 cohorts and 3 subjects per cohort yielded the optimal $\Delta = 0.45$ for the first 2 cohorts and $\Delta = 1.05$ for the cohorts 3-8 (Ivanova and Kim, 2009). Following this suggestion, we propose setting $\Delta_{n_i} = 3/[1+\exp(3-0.05 n_i)]$. Defining Δ_{n_j} in such a way makes Δ equal to about 0.5 for small n_i, equal to about 1.0 when $n_i = 46$; Δ_{n_j} tends to 3.0 when n_i goes to infinity. The choice of the value φ is guided by the optimal

allocation for estimating of the peak dose. The value of φ = 1 results in equal allocation (which is nearly optimal) to the target dose and a dose level right below in the limit, therefore we set φ = 1 in the adaptive strategy. This results in a simple allocation strategy for estimating a target dose when dose-response curve is assumed to be non-decreasing: increase the dose if $T_i \le -\Delta_{n_i}$, where n_i is the number of subjects assigned to dose d_i so far; otherwise decrease the dose.

4.3.2 Covariate adjusted randomization.

In a dose-ranging study subjects are usually assigned in cohorts. In a trial estimating the MED, at each step a dose is adaptively chosen from $\{d_1,...,d_K\}$ and some subjects in a cohort are randomized to placebo d_0. In a trial estimating the peak dose some subjects in each cohort are randomized to d_K and some to one of $\{d_1,...,d_{K-1}\}$. To ensure balanced allocation between d_K and the estimated target dose we propose to keep the allocation to d_K approximately equal to the allocation to a dose with the most assignments. For trials estimating the MED the allocation is kept balanced between placebo and a dose with maximum assignments. In the remainder of this section we will use the estimation of the peak dose as an example. In order to achieve balance in assignments with respect to covariates between d_K and the current dose recommended by the adaptive strategy based on the data available so far, we propose to use a method similar to minimization (Taves 1974; Pocock and Simon 1975). For ease of presentation, we describe the method for a single covariate with two levels $x = 0$ and $x = 1$. Let $n_{ix}(t)$ be the number of subjects assigned to dose d_i, $i = 0,\ldots,K$, with the covariate level x, $x = 0, 1$, right after subject t has been assigned, and let $n_i(t) = n_{i0}(t) + n_{i1}(t)$, that is, $\sum_{i=0}^{K}\sum_{x=0}^{1} n_{ix}(t) = \sum_{i=0}^{K} n_i(t) = t$. Let dose d_i be the

current dose. A new subject, subject $t + 1$, entering the study will be assigned to either d_i or to the highest dose d_K. Define the measure of discrepancy (MD) as follows

$$MD = W\left|n_K(t+1) - \max_{i=1,\ldots,K-1} n_i(t+1)\right| + \left|\frac{n_{Kx}(t+1)}{n_K(t+1)} - \frac{n_{ix}(t+1)}{n_i(t+1)}\right| \frac{n_i(t+1)n_K(t+1)}{n_i(t+1)+n_K(t+1)}$$

Here W is the weight similar to the weight used in minimization; we used $W = 0.5$ in the simulation study. The value of MD is computed assuming that subject $t + 1$ is assigned to d_i, and then computed assuming that subject $t + 1$ is assigned to d_K. The subject is assigned to the dose with the smaller value of MD. In the case when the values of MD are the same, the subject is randomized to one of the doses with equal probability. When there are no covariates, this strategy is still useful as it helps to keep the number of subjects assigned to placebo approximately equal to the number of subjects assigned to the estimated target MED.

4.4 Simulation Study

Our simulation study investigates the effect of balancing assignments to doses with respect to covariates, adjusting for covariates, doing both or neither, and compared adaptive strategies with equal allocation. Simulation results are based on 5000 simulation runs. Table 6 displays seven scenarios from Bretz et al. (2005) that we considered. A two-level covariate, in-patient with $x = 0$, $\Pr(x = 0) = 0.4$, and out-patient with $x = 1$, with covariate effect $\beta = 0.5$ was considered. To estimate the MED defined as the dose with the mean response equal to $\mu_0 + 0.35$ we used the adaptive strategy described in Section 4.3.1; to balance with respect to covariates we used the algorithm described in Section 4.3.2. Unless specified otherwise the simulations were performed with balancing with respect to covariates and adjusting for covariates in the analysis.

It is useful to have a lead-in phase with equal number of assignments to all doses. Such lead-in phase ensures that all doses are tested and provides data for initial estimation of the MED. The estimated MED after the lead-in is used as a starting dose for the adaptive design. Forty two subjects, 26% of the total sample size of 162, were assigned in the lead-in phase, 6 subjects to each dose. It is always desirable to stop the trial for futility if a drug is not beneficial, that is, if the null hypothesis H_0: $\mu_K = \mu_0 + \eta$, $\eta = 0.35$, is rejected in favor of one-sided alternative $\mu_K < \mu_0 + \eta$. Looks for futility were performed at each interim and at the final analysis. Given the goals of a Phase II trial, we suggest setting the probability of rejecting an efficacious drug at 0.05 or lower. The Pocock stopping boundary was used in sequential monitoring to minimize the expected sample size if the treatment is not effective. If the trial was stopped early for futility or futility was established during the final analysis none of the doses was selected as the estimated MED.

Table 7 displays the proportion of trials in which the true MED was selected as the estimated MED. In adaptive design, after lead-in subjects were assigned in 5 cohorts, 24 subjects per cohort. That is, there were 5 interim and one final analysis. Data were generated from scenarios in Table 6 with $\sigma = 0.25$. We report results when balancing with respect to the covariate, adjusting for the covariate, neither or both were performed. The results where the allocation was balanced with respect to covariates but covariates were ignored in the analysis, and where the allocation was not balanced but covariates were used in the analysis were only slightly worse than those for adaptive design with balancing and adjusting. These findings were consistent across various numbers of analyses and for both the MED and the peak dose estimation. Thus one needs to at least either balance assignments with respect to covariates or adjust for covariates in the analysis. It is interesting to note that balancing with

respect to and adjusting for covariates when $\beta \neq 0$ yielded very similar quality of estimation to the case where covariate effect was 0. On the other hand, it is clear from the comparison of the adaptive adjusted design and adaptive design was no balancing or adjustment with respect to covariates that disregarding covariates that are associated with the outcome in design and analysis have negative effect on quality of estimation of the target dose. Compared with equal allocation, adaptive design yields much higher probability of selecting the correct MED.

Table 8 displays results for the average sample size at the estimated target dose. Adaptive design assigns more subjects on average to the estimated target dose, which increases the power of comparisons that involve the estimated target dose. Adaptive design assigns about twice as many subjects to the estimated target dose on average compared to equal allocation in all non-null scenarios. As far as futility stopping, all trials were stopped for futility in the null scenario, scenario 1, and the average sample size in scenario 1 was 63. For the equal allocation, in the null scenario futility was declared at the end of every trial after 160 patients were treated. None of trials was stopped early for futility in scenarios 2-7 when adaptive design was used.

To study the effect of the number of interim analyses on the design performance we performed simulations with the total of 2 analyses (with two doses selected in stage 2), 3 analyses (2 cohorts of size 60), 4 analyses (3 cohorts of size 40), 6 analyses (5 cohorts of size 24), 9 analyses (8 cohorts of size 15) and 13 analyses (12 cohorts of size 10). Each trial had a lead-in phase with 42 subjects total equally allocated to 7 doses. The average (over non-null scenarios) percent selection of the target dose was 0.84, 0.87, 0.88, 0.88, 0.88 and 0.89. The average sample size at the estimated target dose (averaged over seven scenarios) was 38, 39,

44, 44, 43 and 43. Regardless of the number of analysis, the probability of stopping early for futility was 1 in the null scenario, the average sample size was 117, 84, 72, 62, 58 and 58 for 2, 3, 4, 6, 9, and 13 for 2, 3, 4, 6, 9, and 13 analysis correspondingly. None of trials was stopped early for futility in scenarios 2-7.

We also repeated simulations with different values of σ. The probabilities of correct selection of the MED in non-null scenarios were 0.49, 0.37, 0.38, 0.67, 0.65, 0.76 for σ = 0.65. These numbers are to be compared with selection probabilities for equal allocation, 0.39, 0.30, 0.33, 0.58, 0.61, 0.69. Average sample sizes at the estimated MED were 37, 36, 36, 39, 42, 41 for σ = 0.65. The probability of stopping early for futility was 0.6 in scenario 1, with the average sample size of 142. For equal allocation futility was declared in 0.23 of the trials. None of trials was stopped early for futility in scenarios 2-7 when adaptive design was used. When the variability of the outcome was large (σ = 1.3 and larger) while keeping the sample size the same, selection probabilities were low and adaptive design did not bring much benefit compared to equal allocation.

Often patient response is not known prior to assignment of the next cohort. We repeated the simulations under the following staggered entry model: the outcome is available on day 7 after the start of the treatment; the accrual rate was 5 patients per week; the adaptation was performed at the time when new cohort is initiated based on all data available at that point. The probability of correct selection for adaptive design decreased on average by 0.01 for each scenario with the almost same number of patients at each dose.

We also performed simulations to estimate the peak dose defined as the lowest dose with the mean response of $\mu_K - \gamma$, where $\gamma = 0.06$. The conclusions were similar to those of the MED. With lead-in phase allocating 6 subjects per dose and four cohorts with 30 patients

each, the probabilities of correct selection of the peak dose in non-null scenarios were 0.46, 0.70, 0.55, 0.73, 0.88, 0.95. These results should be compared to 0.38, 0.55, 0.47, 0.66, 0.81, 0.88 for equal allocation. There was almost no additional benefit in increasing the number of analyses and slight benefit as far as the average sample size at the target dose. For $\sigma = 0.65$, the probabilities of correct selection of the peak dose were 0.27, 0.35, 0.40, 0.51, 0.56, 0.67 for adaptive design, and 0.22, 0.31, 0.34, 0.43, 0.50, 0.55 for equal allocation. Estimation of the peak dose would be normally one of the objectives of a dose-finding trial with other objectives being, for example, the estimation of the MED and/or establishing dose-response. Such a trial will have a stopping rule for futility described earlier. If futility is established during the trial none of the doses is to be selected as the estimated peak dose.

4.5 Discussion

We proposed sequential strategies to estimate the MED and the peak dose. The strategies are based on the optimal design that maximizes the probability of correctly selecting the target dose. In many trials, it is desirable to test the mean response at the target dose against placebo mean response. In this case the optimality criterion for the design can be set as a function of the probability of correct selection of the target dose and the number of subjects allocated to target. Another possibility is to fix the probability of correct selection at, say, 95% of the optimal and maximize the number of subjects allocated to the target dose.

One of the limitations of the proposed approach is that the dose-response curve is assumed to be non-decreasing. For certain compounds dose-response curve can have a downturn in the high range. In that case one can use umbrella isotonic regression (Roberson, Wright and Dykstra, 1988) in place of isotonic regression. Then an adaptive strategy can be designed to estimate the MED and peak doses.

CHAPTER 5 TWO-STAGE DESIGNS FOR PHASE II DOSE-FINDING TRIALS

We propose Bayesian adaptive two-stage design to efficiently estimate the minimum effective dose or the maximum dose in a dose-finding trial where some monotonicity assumptions regarding dose-response relationship can be made. The new design allocates subjects in stage 2 according to the posterior distribution of the location of the target dose. Simulations show that the proposed two-stage design is superior to equal allocation and to two-stage strategy where only one dose is left in stage 2.

KEY WORDS: Dose ranging; Minimum effective dose; Maximum dose; Phase 2 trials; up-and-down designs.

5.1 Introduction

Phase 2 dose-finding studies are central for drug development as they identify a small subset of doses that are further investigated in a Phase 3 trial. Efficacy is the most important consideration in selecting candidate doses of a drug. The minimum effective dose (MED) is the smallest dose with a discernible useful effect (ICH E4 Guideline, 1994). The MED is often defined as the dose with the mean efficacy outcome equal to a certain target, where target is defined as compared to placebo. Mean efficacy is usually assumed to be non-decreasing with dose. Often times both efficacy and safety are taken into consideration when selecting the best dose, as increasing the dose results in both higher efficacy and increased toxicity or adverse event rates. Common approach is to quantify efficacy and adverse event rate trade-off through a utility function. Such function incorporates both efficacy and safety

into a measure of overall clinical "utility" (Thall, Inoue and Martin, 2002; Ivanova, 2003; Dragalin and Fedorov, 2006; Thall and Cook, 2004; Berry et al., 2001; Fedorov and Wu, 2007; Ivanova et al., 2009). Utility function often has an "umbrella" or "inverse U" shape and the objective of a trial is to maximize overall clinical "utility" of the drug. We will refer to the dose that maximizes the utility function as the optimal dose. An additional objective may be to test efficacy and adverse event rates at the MED or the optimal dose against placebo and/or an active control. Therefore a good assignment strategy for a dose-ranging study will be a strategy that provides good quality of estimation of the target dose and increased sample size at the estimated target dose to yield better power of treatment comparisons.

Most dose-finding designs (one notable exception is the design in Berry et al. 2001) assume a certain order of the means of the dose-response curve, non-decreasing, umbrella or various partial orders. In several recent publications the dose-response curve was successfully estimated under an isotonic model without assuming a parsimonious model for dose-response relationship. Conaway et al. (2004), Yuan and Chappell (2004), Ivanova and Wang (2005) and Ivanova and Kim (2009) proposed frequentist methods; Li et al. (2008) and Bekele et al. (2008) proposed Bayesian approaches for various dose-finding problems.

Most of the dose-finding designs are fully sequential as they have been proposed for oncology studies where experimentation starts at the lowest dose and dose is gradually escalated because of ethical considerations. Perhaps that is why most of the adaptive dose-finding methods for non-oncology dose-finding studies are also fully sequential or multi-stage designs (Berry et al., 2001; Ivanova, Bolognese and Perevozskaya, 2008; Ivanova et al., 2009). The logistics of implementing a fully sequential dose finding study can be daunting.

On the other hand, the most common design in dose-ranging studies is equal allocation to all doses. A two-stage design seems to be a reasonable compromise between a multistage and a single-stage approach. Miller, Guilbaud and Dette (2007) recently investigated a two-stage strategy for a dose-ranging study that is optimal across several parametric models. They concluded that the proposed two-stage strategy offers minor benefit compared to a single-stage design in terms of the efficiency of estimation of the target dose. Dragalin et al. (2008) investigated optimal two-stage designs for two correlated binary endpoints that follow a bivariate probit model and concluded that two-state strategy is superior to equal allocation.

In this paper we propose Bayesian two-stage designs for dose-ranging trials under the following three models: estimating MED under assumption of non-decreasing dose-response curve, estimating the dose with the highest response under umbrella order assumption, estimating MED under isotonic matrix order. The latter problem arises when several different administration schedules are investigated.

5.2 The Model

Let $\{d_1,...,d_K\}$ be the set of ordered dose levels selected for a trial with d_1 denoting placebo. Ignoring the monotonicity, a conjugate prior density (Gelman et al., 2004, p. 78) can be specified as

$\mu_j \mid \sigma^2 \sim N(\mu_{0j}, \sigma^2 / k_{0j})$, j = 1, 2,...,K, and $\sigma^2 \sim IG(\nu_0, \sigma_0^2)$,

where *IG* denotes inverse gamma distribution. Let n_j be the number of subjects assigned to d_j, $n = n_1+\ldots+n_K$. Subjects' response at d_j, $\boldsymbol{y_j}$, j = 1, 2,...,K is a vector of n_j i.i.d. $N(\mu_j, \sigma^2)$ random variables. The joint posterior of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and σ^2 is the product of

$\mu_j \mid \sigma^2$, $y \sim N(M_j; V_j)$, j = 1, 2,...,K, and $\sigma^2 \mid y \sim IG(\nu_n, \sigma_n)$, (1)

where $M_j = (k_{0j}\mu_{0j} + n_j\bar{y}_j)/(k_{0j} + n_j)$, $V_j = \sigma^2/(k_{0j} + n_j)$, $\nu_n = \nu_0 + n/2$, and

$$\sigma_n = \sigma_0 + \frac{1}{2}\sum_{j=1}^{K}\left\{(n_j - 1)s_j^2 + \frac{k_{0j}n_j}{k_{0j} + n_j}(\bar{y}_j - \mu_{0j})^2\right\},$$

with $(\bar{y}_1, \ldots, \bar{y}_K)'$ denoting the unrestricted maximum likelihood estimates, and s_j^2 denoting the empirical variance of $\boldsymbol{y}_j$. We follow the approach of Dunson and Neelon (2003) and Gunn and Dunson (2005) and map unconstrained mean vector $\boldsymbol{\mu}$ from $R^K \to \Omega$ to obtain the posterior distribution for the restricted means. Here $\Omega \subset R^K$ is defined by a set of inequalities on the elements of $\boldsymbol{\mu}$. Since the posterior distribution (1) of unconstrained parameter vector $\boldsymbol{\mu}$ follows a simple conjugate form, we can easily obtain the draws via Gibbs sampling algorithm, and transform draws to the constrained draws from the posterior density for the constrained parameter vector, $\boldsymbol{\mu}^*$, using the isotonic transformation approach. In the following sections we consider three types of constraints that define Ω: non-decreasing, umbrella and matrix order.

5.3 Two-stage Design to Find the MED

5.3.1 Estimating the MED.

In this section dose-response is assumed to be non-decreasing with dose, $\mu_1 \le \ldots \le \mu_K$, and the goal is to find the minimum effective dose MED defined as the dose with the mean response of $\mu_1 + \eta$, where $\eta > 0$ is the minimum clinically important difference specified before the trial. Under the assumption of non-decreasing dose-response $\mu_1 \le \ldots \le \mu_K$, in non-Bayesian set-up, the restricted maximum likelihood estimate for $\boldsymbol{\mu}$, $(\mu_1^*, \ldots, \mu_K^*)'$, can be obtained from the unrestricted maximum likelihood estimates, $(\bar{y}_1, \ldots, \bar{y}_k)'$ (Robertson, Wright and Dykstra, 1988) as:

$$\hat{\mu}_j^* = \min_{t \in U_j} \max_{s \in L_j} \left(\frac{\sum_{h=s}^{t} n_h \bar{y}_h}{\sum_{h=s}^{t} n_h} \right), \qquad (2)$$

for $j = 1, 2,...,K$. Here $L_j = \{1,..., j\}$ and $U_j = \{j,..., K\}$. Transformation (2) is a least-squares projection form R^K to the restricted space Ω. In the Bayesian setting, following Dunson and Neelon (2003) we project draws from the unconstrained posterior density (1) onto Ω using a minimal distance mapping. We then work with transformed draws.

5.3.2 Two stage design.

The two-stage strategy we proposed is described below. Let N_1 and N_2 be the total sample sizes in two stages correspondingly. The question of how to split the total sample size $N_1 + N_2$ between the two stages in the best way is considered in Section 5.6.

Step 1. In stage 1, assign N_1/K subjects to each dose.

Step 2. Update the prior using stage 1 data to obtain unconstrained posterior density of $\boldsymbol{\mu}$. Transform each of D draws from the unconstrained posterior density of $\boldsymbol{\mu}$ to follow non-decreasing order as described in Section 5.3.1. For each draw the location of the MED is determined as the dose with the value closest to $\mu_1 + \eta$. These locations are summarized as the posterior distribution for the location of the MED $\boldsymbol{\pi} = (\pi_1,..., \pi_K)$.

Step 3. Let N_2 be the number of subjects available for stage 2 and $\pi_m = \max_{j=2,...,K}(\pi_j)$. In stage 2, $\pi_j /(1-\pi_1+\pi_m)N_2$ of subjects are assigned do dose $d_j, j = 2,\ldots, K$ and $\pi_m /(1-\pi_1+\pi_m)N_2$ subjects are assigned to placebo. That is, subjects are allocated proportional to the posterior of the MED location except for placebo, where the number of subjects is set to be equal to the number allocated to the most likely target dose.

Step 4. After stage 2 the data from both stages are combined. The estimated MED is the dose d_j , $j = 2,\ldots,K$, such that the posterior mean of $\mu_j - \mu_1$ is the closest to η.

5.3.3 Comparing the MED with placebo.

It is often of interest to compare the target dose with placebo and/or an active control. This comparison should account for both the multiplicity of treatments in stage 2 and the selection processes (i.e. interim analysis). The classical Dunnett's test adjusts for the original number of hypotheses but does not take into account selection process. Our simulations show that when used with our two-stage design the Dunnett's test is conservative in terms of controlling of the family-wise type I error rate for comparing the MED to placebo. These conclusions are similar to those in Koenig et al. (2008). Another approach is to use a combination test with the weighted inverse normal combination function applied together with the closed testing principle (Posch et al., 2005). In the combination test, *t*-test p-values are calculated for each dose and each stage. The closed testing principle with Simes' test of intersection hypotheses is then used within each stage. The overall p-value for each dose is calculated by applying a weighted inverse normal combination function to the two adjusted p-values, p and q,

$$C(p,q) = 1 - \Phi[\sqrt{w}\Phi^{-1}(1-p) + \sqrt{1-w}\Phi^{-1}(1-q)].$$

Here w, $w \geq 0$, is a pre-defined weight, and Φ is the cumulative distribution function of the standard normal distribution. The adaptive combination test performs well when one treatment comparison is made in stage 2, however, due to the closed testing principle, the adaptive combination test becomes more conservative if more arms are selected after the first stage. As our proposed two-stage design can have any number of treatment arms in stage 2, both testing procedures will be conservative. Instead, we propose to simulate the distribution

of the Dunnett's p-values under the null hypothesis following the proposed two-stage design, then the critical value for the test can be obtained as 0.025 percentile of the Dunnett's p-values distribution. Our simulations show that the critical value depends on the number of doses *K*, as well as the variance of the outcome. We recommend using conservative guess for the variance to obtain the critical value. Such approach will preserve the type I error rate and is more efficient than using the Dunnett's test. If there is no knowledge regarding the variance of the outcome, Dunnett's test will always preserve the type I error rate.

5.4 Two-stage Design to Find the MED When Two Administration Schedules are Investigated

Several administration schedules can be studied in a phase 2 trial. Often schedules can be ordered based on intensity, for example, twice a day administration is more intense than once a day, with twice a day yielding higher or same mean response compared to once a day. This leads to two-dimensional monotonicity assumption: 1) mean response is non-decreasing with dose given the schedule, and 2) mean response is non-decreasing with schedule given the dose. Let $\mu_{11},...,\mu_{1K}$ be the vector of mean responses for once a day schedule and $\mu_{21},...,\mu_{2K}$ for twice a day. We have $\mu_{11} \le ... \le \mu_{1K}$, $\mu_{21} \le ... \le \mu_{2K}$, and $\mu_{1i} \le \mu_{2i}$ for any $i = 2,\ldots,K$. As the first dose is placebo, we additionally have $\mu_{11} = \mu_{21}$. This order is often referred to as marix order (Robertson, Wright and Dykstra, 1988). The maximum likelihood estimates under matrix order restriction, μ^*, can be computed using the Dykstra *et al.* algorithm (Robertson, Wright and Dykstra, 1988) that can be found in the Appendix C.

The goal of the trial with two administration schedules can be to estimate one MED or to estimate two MEDs, one for each administration schedule. The two-stage design we

propose will work for either strategy. The two-stage design is similar to the one in Section 5.3. In the final analysis, depending on the objective, either one MED is selected or two MEDs, one for each administration schedule. When comparing the estimated MED with placebo as in Section 5.3, we generated the critical value from the distribution of Dunnett's p-values obtained under the null hypothesis. As in the case of a single administration schedule, the critical value depends on the number of doses as well as on σ^2.

5.5 Two-stage Design to Find the Optimal Dose

In this section we consider a problem of finding the maximum of a utility function. We assume the umbrella order $\mu_1 \le ... \le \mu_{h-1} \le \mu_h \ge \mu_{h+1} \ge ... \ge \mu_K$, where the location of the peak, *h*, is unknown. First, assuming a known peak location *k*, the restricted estimates can be obtained as follows:

$$\mu_j^{*k} = \min_{t \in U_k^k} \max_{s \in L_j^k} \left(\frac{\sum_{h=s}^{t} n_h \bar{y}_h}{\sum_{h=s}^{t} n_h} \right), \qquad (3)$$

for $j = 1, 2,...,K$. Here U_j^k and L_j^k denote subsets of {1, … , *K*} such that the ordering $\mu_{j'} \le \mu_j$ is known for all $j' \in L_j^k$ and the ordering $\mu_{j'} \ge \mu_j$ is known for all $j' \in U_j^k$. To allow for a peak at an unknown location, k, we choose μ^* by minimizing the distance across different choices of peak:

$$\boldsymbol{\mu}^* = \min_{k \in \{1,...,K\}} \{(\boldsymbol{\mu}^{*k} - \boldsymbol{\mu}) \Sigma_{\boldsymbol{\mu}}^{-1} (\boldsymbol{\mu}^{*k} - \boldsymbol{\mu})'\}, \qquad (4)$$

where $\Sigma_{\boldsymbol{\mu}} = diag(V_1,...,V_K)$.

As in Gunn and Dunson (2005), we transform the unrestricted draws using formulae (3) and (4) and then consider the draws of μ^* to be draws from a Bayesian posterior.

The two-stage design for estimating the optimal dose is as follows.

Step 1. In stage 1, assign N_1/K subjects to each dose.

Step 2. Update the prior using stage 1 data to obtain unconstrained posterior density of $\boldsymbol{\mu}$. Transform each of D draws from the unconstrained posterior density of $\boldsymbol{\mu}$ to follow umbrella order to obtain the posterior distribution for the maximum of the umbrella $\boldsymbol{\pi} = (\pi_1, ..., \pi_K)$.

Step 3. Let N_2 be the number of subjects available for stage 2 and $\pi_m = \max_{j=2,...,K}(\pi_j)$. In stage 2, $\pi_j / (1 - \pi_1 + \pi_m) N_2$ of subjects are assigned do dose $d_j, j = 2, \ldots, K$ and $\pi_m / (1 - \pi_1 + \pi_m) N_2$ subjects are assigned to placebo. That is, subjects are allocated proportional to the posterior of the optimal dose location except for placebo, where the number of subjects is set to be equal to the number allocated to the most likely target dose.

Step 4. The optimal dose is estimated from combined stage 1 and 2 data, as the mode of the posterior distribution for the optimal dose location.

Methods similar to that in Section 5.3.3 are used to compare the estimated optimal dose with placebo. The critical value for the test is obtained as 0.025 percentile of the distribution of Dunnett's p-values under the null following a two-stage design to estimate the optimal dose. Simulations show that the critical value depends on the number of doses K, however, unlike the MED case, it does not depend on unknown variance σ^2.

5.6 Simulation Study

We performed a simulation study to compare the performance of the proposed two-stage design with a two stage design where only one treatment arm is selected for the second stage, and with a single stage design with equal allocation. In both of these designs we select the mode of the posterior distribution of the target dose location. In two stage design with a single arm carried in stage 2, this is done based on stage 1 data; in a single stage design - based on all data. For the two-stage design with a single arm carried in stage 2, we used the combination test with $w = 0.5$ with Simes' method for testing. We used Dunnett's test in a single stage design.

Simulation results are based on 10000 simulation runs. The total number of subjects in a trial was 180 to estimate the MED, 252 to estimate the MED with two administration schedules and 100 to estimate the optimal dose. Scenarios 1-5 in Table I were used for the MED simulations and scenarios 6-10 for the optimal dose simulations. All dose-response shapes are from Bretz, Pinheiro and Branson (2005) with doses (0, 0.05, 0.45, 0.8, 1) for MED and doses (0, 0.25, 0.5, 0.75, 1) for the optimal dose simulations. Table 10 displays scenarios with two administration schedules. The dose-response curves are from Bretz, Pinheiro and Branson (2005) with doses (0.05, 0.15, 0.30) and (0.40, 0.70, 1.0) for scenario 1, and doses (0.05, 0.45, 0.85) and (0.30, 0.70, 1.0) for scenario 2. For a scenario with mean vector $\boldsymbol{\mu}$ outcomes at d_j follow normal distribution $N(\mu_j, \sigma^2)$ with $\sigma = 0.65$. The conjugate prior for μ_j follows the conditional distribution $\mu_j \mid \sigma^2 = N(\mu_{0j}, \sigma^2 / k_{0j})$, $j = 1$, 2,...,*K*, and $\sigma^2 \sim IG(a_0, b_0)$, with $\mu_{0j} = 0$, $\sigma^2 = 1$, $k_{0j} = 0.001$, and $a_0 = b_0 = 0.001/2$. We obtain the draws from the posterior density (1) via Gibbs sampling algorithm, and transform draws to the constrained draws from the posterior density of the constrained parameter, $\boldsymbol{\mu}^*$,

as described in Section 5.3.1, 5.4 or 5.5. The above process is repeated 1500 times discarding the first 500 iterations as a burn-in.

First, we investigated what is the optimal way to split the total sample size between the two stages. The best proportion for the optimal dose problem was selected based on average power for placebo – target dose comparison. Figure 3 presents power averaged over corresponding scenarios for the three set-ups. Probability of correct selection follows similar pattern and similar results were observed for other values of σ^2. Allocating 0.58 of the sample size in stage 1 gives the best average power when the MED is estimated under non-decreasing or matrix order, allocating 0.42 of the sample is the best to estimate the optimal dose. In the simulation study we used proportion 0.5 and allocated equal number of patients in stage 1 and stage 2.

Table 11 reports simulation results for the MED estimation. Reported are the probability of selecting each dose as the target dose and probability of correctly rejecting the null hypothesis of equality of placebo mean response with the estimated MED. Adaptive design yields the same probability of correct selection compared to the equal allocation. However, it assigns more subjects to the target dose on average compared to equal allocation which leads to much better power. The number of subjects assigned to the estimated MED is equal to 36 for equal allocation, compared to the median number of 46 for the adaptive two-stage strategy with 36 and 52 being the 25the and 75th percentiles. The new adaptive strategy yields much higher probability of correctly selecting the MED and higher power for comparing with placebo than the two-stage strategy where one dose is left in stage 2. The critical value for the adaptive two-stage design was obtained by simulating 40,000 trials under the null hypothesis using true $\sigma^2 = 0.65$. Though the critical value depends on σ^2, the

critical values obtained for σ^2 in [0.4, 0.9] were almost the same as for $\sigma^2 = 0.65$. Since critical value increases as variance increases, our recommendation is to use low bound of the guess for the variance to obtain the critical value.

We also compared the new two-stage strategy to estimate the MED with multi-stage *t*-statistic design from Ivanova, Bolognese and Perevozskaya (2008). A total 180 subjects were assigned in 20 cohorts of size 9. In the first four cohorts, 5 subjects in each cohort were assigned to placebo to provide good estimate of placebo response early in the trial. After that, 3 subjects in each cohort received placebo. Non-placebo assignments throughout the trial were determined according to design in Ivanova, Bolognese and Perevozskaya, 2008. The total number of subjects assigned to placebo was 68 subjects. Multi stage strategy yielded slightly better probability of selecting the correct MED 0.71, 0.73, 0.47, 0.78 and 0.81 for scenarios 1-5 correspondingly, compared to 0.67, 0.67, 0.47, 0.79 and 0.76 for the Bayesian two-stage design. The number of subjects assigned to the estimated MED was significantly higher: median (25the; 75th percentiles) were 72 (52; 88) for the *t*-statistic design compared to 46 (36; 52) for the two-stage strategy, yielding much higher power of the MED - placebo comparison.

Table 12 shows simulation results for selecting the MED in case of two administration schedules. We simulated trials where the goal was to identify one MED. The adaptive two-stage strategy performs better comparatively than in the case of a single schedule. This is because it utilizes additional isotonic assumptions. The adaptive strategy yields better estimation and significantly better power compared to the other two designs. The number of subjects assigned to the estimated target dose is 36 for equal allocation,

compared to the median number of 47 for the adaptive two-stage strategy with 37 and 56 being the 25the and 75th percentiles.

Table 13 displays results for the optimal dose estimation. The conclusions are very similar to the ones for the MED, except this time both two-stage strategies yield much higher power than equal allocation. The number of subjects assigned to the estimated optimal dose is 20 for equal allocation, compared to the median number of 29 for the adaptive two-stage strategy with 23 and 32 being the 25the and 75th percentiles.

We compare the two-stage strategy to estimate the optimal dose with the multi-stage design from Ivanova et al. (2009). As recommended in Ivanova et al. (2009), we assigned 40% of the total sample size of 100 in stage 1 allocating 40 subjects equally among doses, after that, subjects were assigned in cohorts of 6. One to two subjects in each cohort received placebo and the rest received drug. The number of placebo assignments in each cohort was varied in order to keep the total number of placebo assignment approximately equal to the number of assignments at the best dose. This was done to ensure good power of optimal dose – placebo comparison at the end of the trial. Multi-stage strategy did not improve the likelihood of selecting the optimal dose: the probabilities of correctly selecting the optimal dose were 0.53, 0.82, 0.96, 0.86 and 0.84 for scenarios 6-10 compared to 0.51, 0.80, 0.95, 0.85 and 0.83 in Bayesian two-stage design. The number of subjects assigned to the estimated optimal dose was also similar: median (25the; 75th percentiles) were 29 (28; 30) for the mulit-stage design compared to 29 (23; 32) for the two-stage strategy. Multi-stage design is more efficient compared to the two-stage design when more doses are studied, larger sample size is used or the variability of the outcome σ^2 is smaller.

We investigated three-stage strategies similar to the proposed Bayesian two-stage design. Our conclusion was that adding a stage to the two-stage design does not improve power and selection probability by much.

5.7 Discussion

Adaptive two-stage design is a reasonable alternative to equal allocation and multistage strategies. Compared to a single stage design with equal allocation, it yields larger sample size at the estimated target dose and hence provides better power for treatment comparison. The logistics of a two-stage trial are more complex compared to a single stage design but easier than a muti-stage approach. Two-stage approach allows for an interim analysis after stage 1 to stop the trial for futility or efficacy using Bayesian decisions rules.

Often time there is a set of covariate believed to be associated with response. Since the MED is defined using placebo as reference, when the mean response is modeled with identity link function using a linear model with covariates, the target doses for different levels of covariate coincide. Therefore the proposed two-stage adaptive strategy can be easily extended to the case when adjustment with respect to covariate is needed. Another possible extension is finding the optimal dose when several administration schedules are considered.

CHAPTER 6 CONCLUSION: HOW TO CHOOSE A DESIGN FOR YOUR DOSE-FINDING STUDY?

6.1 The Choice of a Dose-response Method Depends on the Assumptions Regarding Dose-response Curve.

Existing methods for dose-finding Phase II trials can be classified according to assumptions regarding dose-response relationship. If investigators are unwilling to make any assumptions regarding dose-response curve, for example, when down-turn in higher range of doses cannot be ruled out, the only appropriate adaptive method is the method proposed by Berry et al. (2001). It assumes that the dose-response curve is smooth but no other assumptions are required. If the dose-response curve is believed to be non-decreasing with dose and no further assumptions can be made, one can use isotonic methods (Ivanova et al., 2008; Ivanova and Kim, 2009; Xiao and Ivanova, 2011a; Ivanova, Xiao and Tymofyeev, 2011). Though not suitable when dose-response curve plateaus in the range of interest (Xiao and Ivanova, 2011a), some Phase I oncology methods can be used successfully in Phase II trials with strictly increasing dose-response curve if the outcome is binary and the goal of the study is to find a dose with a certain response rate. These method include, for example, group designs (Wetherill, 1963) and the CRM (O'Quigley et al., 1990). A number of methods that rely on specific shape of a dose-response curve have been proposed (Dragalin and Fedorov, 2006; Miller, Guilbaud and Dette, 2007). These methods usually are not robust when assumptions regarding the dose-response curve do not hold. For example, the method of Miller, Guilbaud and Dette (2007) assume that the curve follows one of the three

parsimonious parametric models and performs poorly when the true dose response shape is very different from the three shapes.

6.2 The Choice of a Dose-response Method Depends on the Goal of the Study

Adaptive designs can only be useful when the target dose is defined before the study commences. In early stage dose-finding trials, the goal is to collect information on a wide range of doses and specific estimation objectives are not usually of interest. Equal allocation to all the doses in parallel fashion is the most appropriate design choice in this setting. Under the assumption that dose-response curve is non-decreasing, if one or more target doses are of interest, design choice depends on what are the target doses and what hypotheses are being tested. If a single efficacy endpoint is considered, MED and peak doses are usually of interest. If both efficacy and AE rates are considered, one can usually quantify efficacy-AE trade-off via a utility function. The goal is then to find the dose that maximizes the utility function, the optimal dose.

When choosing the most appropriate design, one important component is the number of interim analysis in the study. Increasing the number of interim analysis usually leads to increased sample size at the estimated target dose; however the increase in the precision of estimation of the target dose, measured as proportion of trials where the target dose was selected correctly, is usually modest. As performing multiple interim analyses might be logistically challenging, potential gains in efficiency of estimation of target doses and in power from additional analyses should be evaluated.

Table 14 summarizes isotonic dose-finding methods available for Phase II trials and presents design comparisons.

If there is a need to find the shape of dose-response curve or test for trend, the equal allocation is the best choice. If there are known covariates that might influence the outcome one needs to either balance assignments to doses with respect to covariates or adjust for them in the analysis (Xiao and Ivanova, 2011b).

6.3 Limitations of Adaptive Dose-finding Approaches Proposed in This Dissertation

Below is the summary of limitations of the proposed approaches:

1) Adaptive designs can be used if response is observed relatively quickly compared to the rate of accrual.

2) To use an adaptive design the target dose(s) should be defined before the trial.

3) Adaptive dose-finding designs most likely will not be more efficient than equal allocation, if the goal is to estimate more than two target doses, if it is of interest to assess the shape of a dose-response curve, or if available sample size does not provide adequate precision of mean responses.

4) Proposed methods will yield good results if the true dose-response curve follows the specified order restriction (non decreasing, strictly increasing, or umbrella shape dose response).

APPENDICES

A. Proof of Theorem

Let α_j, β_j, and γ_j denote the probabilities to decrease the dose from d_j to d_{j-1}, to repeat the dose d_j, or increase the dose from d_j to d_{j+1} in $UD(s, c_L, c_U)$. Here $\alpha_j + \beta_j + \gamma_j = 1$ for $j \in \{1,\ldots,K\}$ are the elements of jth row of transition matrix P, with β_j being a diagonal element, and α_j, and γ_j being to the left and to the right of β_j. These probabilities can be computed as follows

$$\alpha_1 = 0,\ \beta_1 = 1-\gamma_1,\ \gamma_1 = \Pr\{Bin(s, p_1) \le c_L\},$$

$$\alpha_j = \Pr\{Bin(s, p_j) \ge c_U\},\ \beta_j = \Pr\{c_L < Bin(s, p_j) < c_U\},\ \gamma_j = \Pr\{Bin(s, p_j) \le c_L\},$$

$$\alpha_K = \Pr\{Bin(s, p_K) \ge c_U\},\ \beta_K = 1-\alpha_K,\ \gamma_K = 0,$$

where $j \in \{2,\ldots,K\text{-}1\}$. The stationary distribution $\boldsymbol{\pi} = (\pi_1,...,\pi_K)$ can be obtained by solving the balance equations, $\pi_j = \pi_{j-1}\,\gamma_{j-1} + \pi_j\,\beta_j + \pi_{j+1}\,\alpha_{j+1}, j \in \{1, \ldots, K\}$ (here for convenience $\gamma_0 = \alpha_{K+1} = 0$). The solution is

$$\pi_j = \prod_{i=1}^{j} \lambda_i, \quad \lambda_1 = \left(1 + \sum_{j=2}^{K} \prod_{i=2}^{j} \lambda_i\right)^{-1}, \quad \lambda_i = \frac{\gamma_{i-1}}{\alpha_i},$$

where $j \in \{2,\ldots,K\}$. Gezmu and Flournoy (2006) showed that γ_j decreases with j while α_j increases with j, so similarly to Durham and Flournoy (1994), the stationary distribution is log-concave, also the mode spans d_{k-1} and d_k if $\lambda_k = 1$. Since $p_1 < p_2 < \ldots < p_j = \ldots = p_K = \Gamma^*$

and Γ^* is a solution of equation (1), $\gamma_i = \alpha_i = \gamma$ for all $i = j,\ldots,K$. Hence

$\lambda_i = \gamma_{i-1} / \alpha_i = \gamma / \gamma = 1$ for $i = j+1, \ldots, K$, and the mode spans doses $d_{(j+1)-1}, \ldots, d_K$.

B. Proof of the Proposition

The vector of unrestricted MLEs obtained from model (1), $\hat{\boldsymbol{\mu}}^U = \left(\mu_0^U, \ldots, \mu_K^U\right)$, has multivariate normal distribution with mean vector $\boldsymbol{\mu} = \left(\mu_0, \ldots, \mu_K\right)$ and variance covariance matrix Σ with $\left(\sigma^2 / n_0, \ldots, \sigma^2 / n_K\right)$ on the diagonal. Let $w_i = n_i / n$, $0 < w_i < 1$, $w_0 + \ldots + w_K = 1$. First we wish to show that when $\mu_s = 0$ and it is known, the probability of correctly selecting the target dose by applying the closes dose or the lowest dose estimator to the weighted average of components of $\hat{\boldsymbol{\mu}}^U$ depends only on vectors $\{w_i\}$, $\left\{\mu_i \sqrt{n} / \sigma\right\}$ and matrix $\left(n / \sigma^2\right)\Sigma$. We note that the probability of selecting the correct target dose can be expressed as:

$$\sum_{j=1}^{J} \int_{A_j} \frac{1}{(2\pi)^{(K+1)/2} \mid \Sigma \mid^{0.5}} e^{-(x-\mu)' \Sigma^{-1} (x-\mu)/2} \prod_{i=1}^{K+1} dx_i \,, \tag{3}$$

where the regions A_j, $j = 1, \ldots, J$, are disjoint sets in the sample space each defined by a set of inequalities where the algorithm chooses the correct target dose. For example, when $\hat{\boldsymbol{\mu}} = \left(\mu_{\tau-1}^U, \hat{\mu}_\tau, \mu_K^U\right)$, the two rejoins are shown in formula (2). We note that each of these regions may be expressed as an intersection of solution sets of inequalities where each side of the inequality is a linear combination of the components of $\hat{\boldsymbol{\mu}}^U$ or the absolute value of such a combination, and the coefficients are functions of the $\{w_i\}$. Thus, the regions A_j of the

sample space where the closest dose is chosen are given by inequalities of the form described above. We rewrite (3) after making the substitution $y_i = \left(x_i\sqrt{n}\right)/\sigma$, $i = 0,\ldots, K$.

$$\sum_{j=1}^{J}\int_{A_j} \frac{1}{(2\pi\sigma^2/n)^{k/2}\,|\,\Lambda\,|^{0.5}} e^{-0.5(y-\frac{\sqrt{n}}{\sigma}\mu)'\Lambda^{-1}(y-\frac{\sqrt{n}}{\sigma}\mu)} \prod_{i=1}^{k}\frac{\sigma}{\sqrt{n}}\,dy_i$$
$$= \sum_{j=1}^{J}\int_{A_j} \frac{1}{(2\pi)^{k/2}\,|\,\Lambda\,|^{0.5}} e^{-0.5(y-\frac{\sqrt{n}}{\sigma}\mu)'\Lambda^{-1}(y-\frac{\sqrt{n}}{\sigma}\mu)} \prod_{i=1}^{k} dy_i,$$

where $\Lambda = \left(n/\sigma^2\right)\Sigma$ is the covariance matrix for the y_i. This substitution does not change the regions A_j because substituting $\left(y_i\sigma\right)/\sqrt{n}$ for x_i in the original inequalities that define the A_j and dividing through by the positive constant $\sigma/\sqrt{n}$ does not change solution set for the inequalities, which now depend only on the w_i and the y_i. Because the y_i will be integrated out, the integral in (3) depends only on the $\{w_i\}$, $\left\{\mu_i\sqrt{n}/\sigma\right\}$ and $\left(n/\sigma^2\right)\Sigma$. If μ_s in the definition of the target dose is not known and is being estimated, the result is obtained similarly to the above by considering $\left(\hat{\mu}_0^{I} - \mu_s^{U},\ldots,\hat{\mu}_K^{I} - \mu_s^{U}\right)$. This is because the density is from a location-scale family, and the regions of correct selection do not depend on the location. Location is irrelevant to the regions of correct selection because the inequalities that define them are comparisons of isotonic estimates where the coefficients for each observation sum to 1, so shifting all means up by a constant does not change their solution sets.

C. The Algorithm for Computing Maximum Likelihood Estimates under Matrix Order (Dykstra and Robertson, 1982)

Step 1. Let $\hat{\mu}^{(1)} = (\hat{\mu}_{ij}^{(1)})$ denote the isotonic regression of μ over rows, i.e. $\hat{\mu}^{(1)}$ minimizes $\sum_{i=1}^{2}\sum_{j=1}^{K}(\mu_{ij} - f_{ij})^2 n_{ij} + (\mu_{10} - f_{10})^2 n_{10}$ subject to $f_{10} \le f_{1j} \le f_{2j}$ for j =1, …, K. Let $R^{(1)} = (r_{ij}^{(1)}) = (\hat{\mu}_{ij}^{(1)} - \mu_{ij})$ be the first set of 'row increments'.

Step 2. Let $\tilde{\mu}^{(1)} = (\tilde{\mu}_{ij}^{(1)})$ denote the isotonic regression over columns of of $\mu + R^{(1)}$, i.e. $\tilde{\mu}^{(1)}$ minimizes $\sum_{i=1}^{2}\sum_{j=1}^{K}(\mu_{ij} + r_{ij}^{(1)} - f_{ij})^2 n_{ij} + (\mu_{10} + r_{10}^{(1)} - f_{10})^2 n_{10}$ subject to $f_{10} \le f_{i1} \le ... \le f_{iK}$ for i = 1,2. Call $C^{(1)} = \tilde{\mu}^{(1)} - (\mu + R^{(1)})$ the first set of 'column increments'. Note that $\tilde{\mu}^{(1)} = \mu + R^{(1)} + C^{(1)}$

Step 3. At the beginning of the mth cycle, $\hat{\mu}^{(m)}$ is obtained by isotonizing $\mu + C^{(m-1)}$ over rows. The mth set of row increments is defined by $R^{(m)} = \hat{\mu}^{(m)} - (\mu + C^{(m-1)})$, so that $\hat{\mu}^{(m)} = \mu + C^{(m-1)} + R^{(m)}$. Next obtain $\tilde{\mu}^{(m)}$ by isotonizing $\mu + R^{(m)}$ over columns. The mth set of column increments is given by $C^{(m)} = \tilde{\mu}^{(m)} - (\mu + R^{(m)})$ or, equivalently, $\tilde{\mu}^{(m)} = \mu + R^{(m)} + C^{(m)}$.

Table 1: Mean response scenarios with their toxicity response curves

Scenario number	Mean response curve	Mean toxicity curve
1	(0.6,0.6,0.6,0.6,0.6,0.6,0.6)	(0.1,0.3,0.4,0.5,0.6,0.7,0.8)
2	(0.3,0.6,0.6,0.6,0.6,0.6,0.6)	(0.1,0.1,0.3,0.4,0.5,0.6,0.7)
3	(0.3,0.3,0.3,0.6,0.6,0.6,0.6)	(0.1,0.1,0.1,0.1,0.3,0.4,0.5)
4	(0.3,0.3,0.3,0.3,0.6,0.6,0.6)	(0.1,0.1,0.1,0.1,0.1,0.3,0.4)
5	(0.3,0.3,0.4,0.5,0.6,0.6,0.6)	(0.1,0.1,0.1,0.1,0.1,0.3,0.4)
6	(0.3,0.4,0.5,0.6,0.7,0.8,0.9)	(0.1,0.1,0.1,0.1,0.4,0.5,0.6)

Table 2: Proportion of trials in which a dose was selected as the expected target dose

Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7
Scenario 1							
Group Design[a]	**0.29**	0.15	0.12	0.11	0.10	0.10	0.13
CRM	**0.33**	0.25	0.21	0.13	0.07	0.02	0.00
t-statistic design	**0.63**	0.22	0.09	0.04	0.01	0.01	0.00
Scenario 2							
Group Design[a]	0.00	**0.32**	0.16	0.13	0.11	0.12	0.15
CRM	0.01	**0.27**	0.28	0.20	0.17	0.06	0.01
t-statistic design	0.02	**0.65**	0.20	0.09	0.03	0.01	0.01
Scenario 3							
Group Design[a]	0.00	0.00	0.01	**0.36**	0.19	0.18	0.26
CRM	0.00	0.00	0.02	**0.33**	0.34	0.24	0.07
t-statistic design	0.00	0.00	0.01	**0.66**	0.21	0.08	0.04
Scenario 4							
Group Design[a]	0.00	0.00	0.00	0.01	**0.41**	0.25	0.34
CRM	0.00	0.00	0.00	0.03	**0.49**	0.34	0.15
t-statistic design	0.00	0.00	0.00	0.01	**0.67**	0.20	0.12
Scenario 5							
Group Design[a]	0.00	0.00	0.01	0.18	**0.29**	0.22	0.29
CRM	0.00	0.00	0.00	0.18	**0.49**	0.25	0.08
t-statistic design	0.00	0.00	0.02	0.30	**0.45**	0.16	0.07
Scenario 6							
Group Design[a]	0.00	0.01	0.22	**0.54**	0.21	0.02	0.00
CRM	0.00	0.00	0.18	**0.61**	0.20	0.00	0.00
t-statistic design	0.00	0.01	0.28	**0.57**	0.13	0.00	0.00

[a] Group design is *UD(4,2,3)*.

Table 3: Average number of patients over trials at each dose

Scenario	d_1	d_2	d_3	d_4	d_5	d_6	d_7
Scenario 1							
Group Design[a]	**22**	17	13	10	7	6	5
CRM	**30**	21	15	9	4	1	0
t-statistic design	**56**	16	5	2	0	0	0
Scenario 2							
Group Design[a]	14	**19**	15	11	8	7	6
CRM	6	**22**	21	16	11	3	0
t-statistic design	14	**45**	14	5	2	1	0
Scenario 3							
Group Design[a]	5	6	13	**19**	14	12	11
CRM	4	4	7	**23**	23	14	4
t-statistic design	7	7	13	**37**	11	3	1
Scenario 4							
Group Design[a]	5	5	6	14	**19**	16	15
CRM	4	4	5	7	**31**	21	8
t-statistic design	7	7	7	12	**34**	10	4
Scenario 5							
Group Design[a]	5	6	10	16	**16**	14	12
CRM	4	4	6	17	**29**	15	4
t-statistic design	7	7	13	24	**21**	6	2
Scenario 6							
Group Design[a]	7	12	19	**22**	14	6	1
CRM	4	6	16	**35**	17	2	0
t-statistic design	7	13	26	**27**	7	0	0

[a] Group design is *UD(4,2,3)*.

Table 4: The distribution of trials at the estimated target dose

Designs	Min	1st	Median	Mean	3th	Max
Scenario 1						
Group Design[a]	4	16	24	25	32	68
CRM	4	40	56	54	68	80
t-statistic design	4	52	80	65	80	80
Scenario 2						
Group Design[a]	4	16	24	23	28	52
CRM	4	36	52	48	60	76
t-statistic design	4	40	56	53	68	76
Scenario 3						
Group Design[a]	4	16	24	23	28	52
CRM	4	36	52	47	60	68
t-statistic design	4	36	44	44	56	68
Scenario 4						
Group Design[a]	4	20	24	24	28	48
CRM	4	40	52	47	60	64
t-statistic design	4	32	40	40	52	64
Scenario 5						
Group Design[a]	4	16	20	22	28	48
CRM	4	32	48	44	56	68
t-statistic design	4	24	36	36	48	72
Scenario 6						
Group Design[a]	4	20	24	24	28	40
CRM	4	32	44	42	56	72
t-statistic design	4	28	40	40	52	76

[a] Group design is *UD(4,2,3)*.

Table 5: Proportion of trials where the estimated target dose is shown to have an efficacy rate significantly better than placebo rate and toxicity rate significantly lower than 0.3

Scenarios	Group design	CRM	*t*-statistic design
Scenario 1	0.53	0.76	0.85
Scenario 2	0.56	0.72	0.84
Scenario 3	0.63	0.76	0.83
Scenario 4	0.64	0.79	0.80
Scenario 5	0.59	0.76	0.70
Scenario 6	0.66	0.76	0.72

Table 6: Data generating dose response curves, d = (0,0.05,0.2,0.4,0.6,0.8,1)

Scenario	Model	Mean Response
1	Constant = 0.6	(0.60,0.60,0.60,0.60,0.60,0.60,0.60)
2	E_{max} = 0.2+0.7d/(0.2+d)	(0.20,0.34,0.55,0.67,0.72,0.76,0.78)
3	Linear in log-dose = 0.2+0.6log(5d+1)/log(6)	(0.20,0.27,0.43,0.57,0.66,0.74,0.80)
4	Linear = 0.2+0.6d	(0.20,0.23,0.32,0.44,0.56,0.68,0.80)
5	Logistic = 0.193+0.607/{1+exp[10log(3)(0.4-d)]}	(0.20,0.21,0.25,0.50,0.74,0.79,0.80)
6	Step 1 = 0.2+0.6I(d≥0.2)	(0.20,0.20,0.80,0.80,0.80,0.80,0.80)
7	Step 2 = 0.2+0.3I(d≥0.4)+0.3I(d≥0.6)	(0.20,0.20,0.20,0.50,0.80,0.80,0.80)

Table 7: Proportion of trials in which the true MED was selected as the estimated MED 1) in adaptive trial with balancing with respect to a covariate and adjusting for covariate in the analysis (Balance and Adjust), in adaptive trial with only balancing (Balancing only), in adaptive trial with only adjusting (Adjusting only), in adaptive trial without balancing or adjusting for covariate (None), in adaptive trial with no covariate effect ($\beta = 0$) and for equal allocation with balancing and adjusting for covariate.

Scenario	*Adaptive, Balancing and Adjusting* $\beta = 0.5$	*Adaptive, Balancing only* $\beta = 0.5$	*Adaptive, Adjusting only* $\beta = 0.5$	*Adaptive, None* $\beta = 0.5$	*Adaptive, None* $\beta = 0$	*Equal allocation, Balancing and Adjusting* $\beta = 0.5$
2	0.88	0.88	0.87	0.76	0.89	0.76
3	0.78	0.73	0.76	0.62	0.78	0.65
4	0.78	0.73	0.78	0.65	0.78	0.69
5	0.96	0.95	0.95	0.88	0.96	0.94
6	0.91	0.89	0.91	0.83	0.91	0.84
7	0.99	0.99	0.99	0.96	0.99	0.99

Table 8: Average number of subjects allocated to the estimated dose in each design. For comparison, equal allocation design with the same total sample size will have 23 subjects at each dose

Scenario	*Adaptive, Balancing and Adjusting* $\beta = 0.5$	*Adaptive, Balancing only* $\beta = 0.5$	*Adaptive, Adjusting only* $\beta = 0.5$	*Adaptive, None* $\beta = 0.5$	*Adaptive, None* $\beta = 0$
2	43	42	38	39	42
3	42	41	35	37	40
4	42	41	36	38	41
5	46	45	44	44	47
6	46	46	45	45	47
7	47	46	46	46	48

Table 9: Dose-response scenarios. Scenarios 1-5 are to illustrate the MED estimation, scenarios 6-10 the optimal dose estimation. The MED and the optimal dose are shown in bold. Placebo is dose one.

Scenario	Model	Mean Response
1	Emax	(0.20,0.34,**0.68**,0.76,0.78)
2	linear in log-dose	(0.20,0.27,**0.59**,0.74,0.80)
3	Linear	(0.20,0.23,0.47,**0.68**,0.80)
4	Truncated-logistic	(0.20,0.20,0.22,**0.54**,0.80)
5	Logistic	(0.20,0.21,**0.58**,0.79,0.80)
6	Quadratic	(0.20,0.60,**0.79**,0.75,0.50)
7	Double-logistic	(0.20,0.37,**0.79**,0.59,0.50)
8	Exponential	(0.20,0.22,0.29,0.43,**0.80**)
9	S5	(0.20,0.50,0.50,**0.80**,0.50)
10	S6	(0.20,0.40,**0.80**,0.60,0.40)

Table 10: Scenarios for the two administration schedules. The MEDs are in bold. Placebo is dose one.

Scenario	Model	Group	Mean Response
1	Linear	A	(0.20,0.23,0.29,0.38)
		B	(0.20,0.44,**0.62**,0.80)
2	Logistic	A	(0.20,0.21,**0.58**,0.80)
		B	(0.20,0.34,0.78,0.80)

Table 11: Probability of correct identification of the MED, and probability of correctly rejecting the null hypothesis that response at the MED is equal to placebo response (Power). The best results are in bold.

Scenario	Design	d_2	d_3	d_4	d_5	Power %
Emax	Two-stage adaptive	0.19	**0.67**	0.1	0.04	**89**
	Two stage, 1 best selected	0.29	0.46	0.14	0.11	84
	Equal allocation	0.22	0.6	0.11	0.06	83
Linear in log-dose	Two-stage adaptive	0.07	**0.67**	0.22	0.04	**90**
	Two stage, 1 best selected	0.18	0.51	0.21	0.1	84
	Equal allocation	0.1	0.62	0.22	0.05	84
Linear	Two-stage adaptive	0.02	0.44	**0.47**	0.07	**87**
	Two stage, 1 best selected	0.08	0.43	0.36	0.12	83
	Equal allocation	0.03	0.45	0.45	0.08	80
Truncated-logistic	Two-stage adaptive	0	0.04	**0.79**	0.17	**86**
	Two stage, 1 best selected	0.02	0.15	0.64	0.19	72
	Equal allocation	0.06	0.06	**0.79**	0.16	76
Logistic	Two-stage adaptive	0.04	**0.76**	0.18	0.02	**90**
	Two stage, 1 best selected	0.12	0.62	0.19	0.06	84
	Equal allocation	0.06	**0.76**	0.17	0.02	82

Table 12: Probability of selecting each dose as the MED when two administration schedules are considered, and probability of correctly rejecting the null hypothesis that response at the estimated MED is equal to placebo response (Power). The best results are in bold.

Scenario	Design	Group	d_2	d_3	d_4	Power %
1	Two-stage adaptive	A	0	0.01	0.15	**89**
		B	0.24	**0.52**	0.05	
	Two stage, 1 best selected	A	0.02	0.05	0.22	73
		B	0.24	0.35	0.13	
	Equal allocation	A	0	0.02	0.17	78
		B	0.23	0.47	0.1	
2	Two-stage adaptive	A	0	**0.64**	0.11	**91**
		B	0.13	0.11	0	
	Two stage, 1 best selected	A	0.03	0.44	0.13	81
		B	0.25	0.12	0.02	
	Equal allocation	A	0.01	0.6	0.1	82
		B	0.17	0.11	0.01	

Table 13. Probability of selecting each dose as the optimal dose, and probability of correctly rejecting at least one null hypothesis (Power). The best results are in bold.

Scenario	Design	d_2	d_3	d_4	d_5	Power %
Quadratic	Two-stage adaptive	0.1	**0.51**	0.38	0.03	89
	Two stage, 1 best selected	0.13	0.45	0.36	0.06	**90**
	Equal allocation	0.09	0.5	0.38	0.02	80
Double-logistic	Two-stage adaptive	0.01	**0.8**	0.14	0.06	**83**
	Two stage, 1 best selected	0.04	0.66	0.2	0.1	83
	Equal allocation	0.01	**0.8**	0.14	0.05	73
Exponential	Two-stage adaptive	0	0.01	0.04	**0.95**	**87**
	Two stage, 1 best selected	0.02	0.03	0.09	0.86	83
	Equal allocation	0	0.01	0.04	**0.95**	77
Step 1	Two-stage adaptive	0.06	0.06	0.83	0.06	82
	Two stage, 1 best selected	0.11	0.1	0.69	0.1	**83**
	Equal allocation	0.05	0.05	**0.85**	0.05	71
Step 2	Two-stage adaptive	0.02	**0.83**	0.16	0.02	**85**
	Two stage, 1 best selected	0.05	0.69	0.21	0.05	84
	Equal allocation	0.01	0.82	0.16	0.01	73

Table 14: Isotonic dose-finding methods available for Phase II trials and design comparisons.

		Number of adaptations				
Target Dose(s)	Testing	1		2		Many
MED	No	Equal	=	IXT2011	<	IK2009
MED	Yes	Equal	<	IXT2011	<	IK2009
Two MEDs	No	Equal	?	IXT2011	?	IBP2008
Two MEDs	Yes	Equal	?	IXT2011	?	IBP2008
PEAK	No	Equal		No design	<	XI2011
Optimal dose	No	Equal	=	IXT2011	<	ILSS2009
Optimal dose	Yes	Equal	<	IXT2011	=	ILSS2009
>2 target doses	Yes/No	Equal				

*Comparison have not been made

IXT2011 = Ivanova, Xiao, Tymofyeev (2011);

IK2009 = Ivanova and Kim (2009);

IBP2009 = Ivanova, A., Bolognese, J., and Perevozskaya, I. (2008);

ILSS2009 = Ivanova, A., Liu, K., Snyder, E., and Snavely, D. (2009);

XI2011=Xiao and Ivanova (2011b)

Figure 1: Optimal allocation to estimate the peak dose. The solid line is the proportion assigned to the true peak dose, d_τ, the dotted line proportion assigned to d_K and the dashed line proportion assigned to $d_{\tau-1}$.

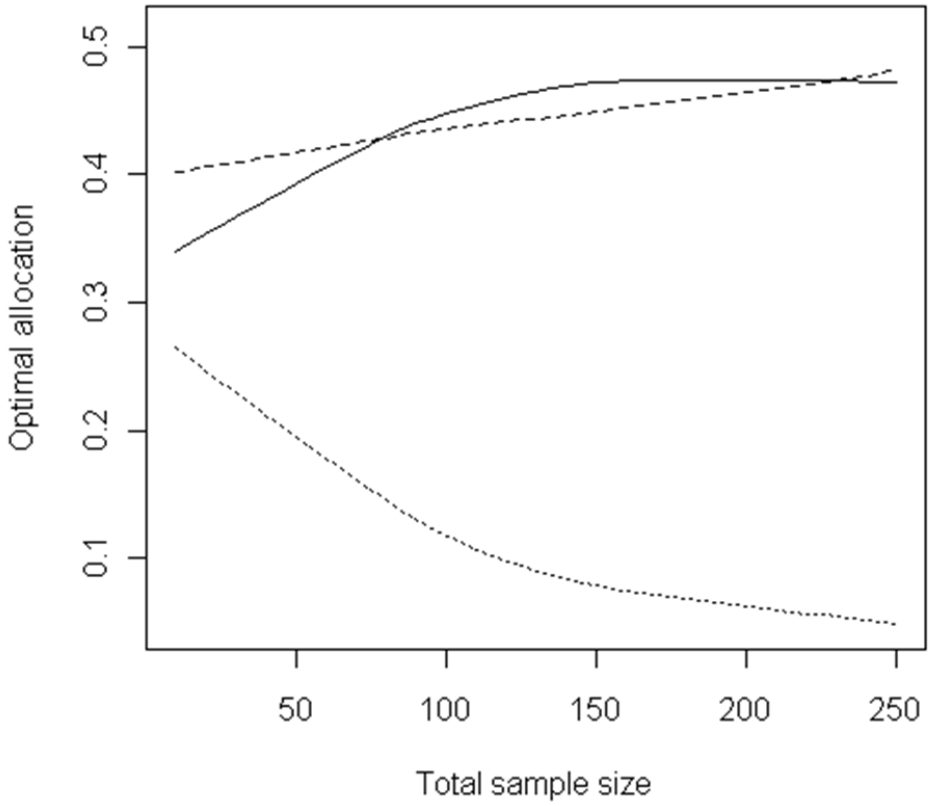


Figure 2: Optimal allocation to estimate the MED. The solid line is the proportion assigned to the true MED, d_τ, the dotted line proportion assigned to placebo d_0, the dashed line proportion assigned to $d_{\tau-1}$ and the dotted-dashed line to $d_{\tau+1}$

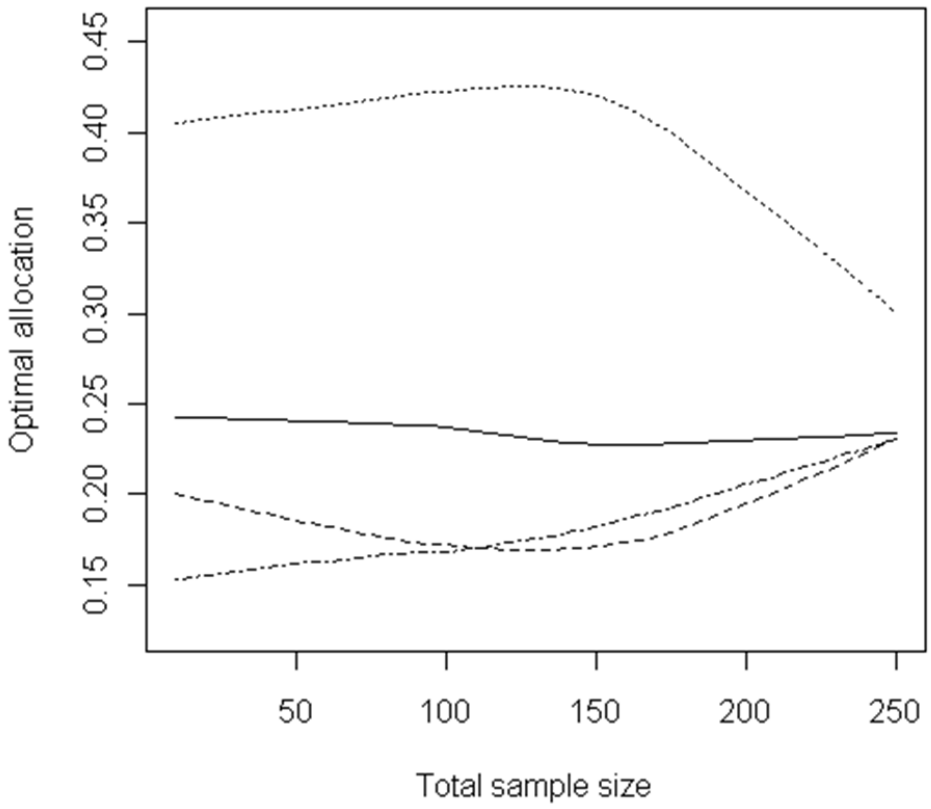
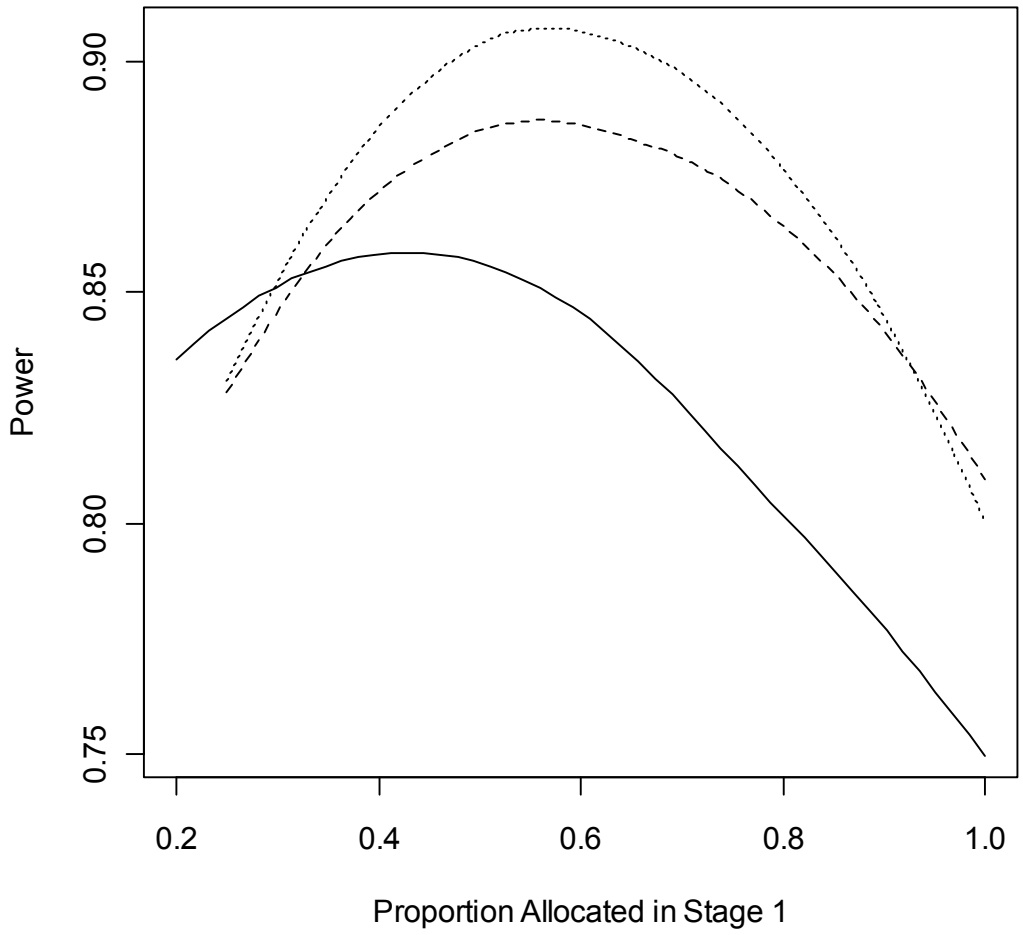


Figure 3: Power averaged over all scenarios plotted against the proportion allocated in stage 1 with proportion of 1.0 corresponding to a single stage design. Solid line corresponds to the optimal dose estimation, dashed line to the MED estimation and dotted line to the MED estimation with two administration schedules.



REFERENCES

1. Atkinson, A.C. (1999). Optimum biased-coin designs for sequential treatment allocation with covariate information with discussion. *Statistics in Medicine* 18, 1741-1755.

2. Babb, J., Rogatko, A., and Zacks, S. (1998). Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 17, 1103-1120.

3. Barlow R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*. John Wiley & Sons, London, New York.

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

5. Berry, D.A., Müller, P., Grieve, A.P., Smith, M., Parke, T., Blazek, R., Mitchard, N. and Krams, M.. Adaptive Bayesian designs for dose-ranging drug trials. In Gatsonis, C., Kass, R.E., Carlin, B., Carriquiry, A., Gelman, A., Verdinelli, I., West, M. (eds), *Case Studies in Bayesian Statistics V* 2001, New York: Springer-Verlag, pp. 99-181.

6. Bretz, F., Pinheiro, J.C., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-Response studies. *Biometrics* 61, 738-748.

7. Cheung, Y.K. (2008). Simple sequential boundaries for treatment selection in multi-armed randomized clinical trials with a control. *Biometrics* 64, 940-949.

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

9. Conaway, M.R., Dunbar S., and Peddada S.D (2004). Designs for single- and multiple-agents phase I trials. *Biometrics* 60, 661-669.

10. Dragalin, V. and Fedorov, V. (2006). Adaptive designs for dose-finding based on efficacy–toxicity response. *Journal of Statistical Planning and Inference* 136, 1800-1823.

11. Dunson D. and Neelon B. (2003). Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics* 59, 286–295.

12. Durham, S.D., Flournoy, N. (1994). Random walks for quantile estimation. Statistical Decision Theory and Related Topics V, Berger, J. and Gupta, S., eds. Springer-Verlag, New York, 467-476.

13. Dykstra, R.L. and Robertson, T. (1982). An Algorithm for isotonic regression for two or more independent variables. *Annals of Statistics* 10, 708-716.

14. Eichhorn, B.H. and Zacks, S. (1973). Sequential search of an optimal dosage, I. *Journal of the American Statistical Association* 68, 594-598.

15. Fedorov, V. and Wu, Y.. Generalized probit model in design of dose finding experiments. In Lopez-Fidalgo, J., Rodriguez-Diaz, J. M., Torsney, B. (eds), *MODA8-Advances in Model Oriented Design and Analysis* 2007, Springer-Verlag, pp 67-73.

16. Gelman A, Carlin J.B., Stern H.S., Rubin D.B.. Bayesian Data Analysis. London: Chapman and Hall, 1995.

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

18. Gunn L. and Dunson D. (2005). A transformation approach for incorporating monotone or unimodal constraints. *Biostatistics* 6, 3, pp. 434-449.

19. Hall, D.B., Meier, U. and Diener, H.C. (2005). A group sequential adaptive treatment assignment design for proof of concept and dose selection in headache trials. *Contemporary Clinic Trials* 26, 349-364.

20. Hwang, J.T.G. and Peddada, S.D. (1994). Confidence interval estimation subject to order restrictions. *Annals of Statistics* 22, 67-93.

21. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH harmonised tripartite guideline: E4 dose-response information to support drug registration. Current step 4 version, March 10, 1994. Available at: http://www.ich.org/LOB/media/MEDIA480.pdf

22. Ivanova, A. (2003). A new dose-finding design for bivariate outcomes. *Biometrics* 59, 1003-1009.

23. Ivanova, A. (2006). Escalation, up-and-down and A+B designs for dose-finding trials. *Statistics in Medicine*, 25, 3668-3678.

24. Ivanova, A., Bolognese, J. and Perevozskaya, I. (2008). Adaptive design based on *t*-statistic for dose-response trials. *Statistics in Medicine* 27, 1581-1592.

25. Ivanova, A., Flournoy, N., and Chung, Y. (2007). Cumulative cohort design for dose-finding. *Journal of Statistical Planning and Inference* 137, 2316-2317.

26. Ivanova, A. and Kim, S. (2009). Dose-finding for binary ordinal and continuous outcomes with monotone objective function: a unified approach. *Biometrics* 65, 307-315.

27. Ivanova, A., Liu, K., Snyder, E. and Snavely, D. (2009). An adaptive design for identifying the dose with the best efficacy/tolerability profile with application to a crossover dose finding study. *Statistics in Medicine* 28(24), 2941-51.

28. Ivanova, A. and Wang, K. (2006). Bivariate isotonic design for dose-finding with ordered groups. *Statistics in Medicine* 25, 2018-2026.

29. Ivanova A., Xiao, C. and Tymofeyev, Y. (2011). Bayesian two-stage designs for phase II

dose-finding trials. In preparation.

30. Koenig F., Brannath W., Bretz F. and Posch M. (2008). Adaptive Dunnett tests for treatment selection. *Statistics in Medicine* 27, 1612–1625.

31. Li, Y., Bekele, B.N., Ji, Y. and Cook, D.C. (2008). Dose-schedule finding in phase I/II clinical trials using a Bayesian isotonic transformation. *Statistics in Medicine* 27, 4895-4913.

32. McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall: London.

33. Miller, F., Guilbaud, O. and Dette, H. (2007). Optimal Designs for Estimating the Interesting Part of a Dose-Effect Curve. *Journal of Biopharmaceutical Statistics* 17, 1097 – 1115.

34. Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal* 7, 308-313

35. O'Quigley, J. and Paoletti, X. (2003). Continual reassessment method for ordered groups. *Biometrics* 59, 430-440.

36. O'Quigley J., Pepe M., and Fisher L. (1990). Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics* 46, 33-48.

37. O'Quigley, J. and Shen L. Z. (1996). Continual reassessment method: a likelihood approach. *Biometrics* 52, 163-174.

38. Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trials. *Biometrics* 31, 103-115.

39. Posch M., Koenig F., Branson M., Brannath W., Dunger-Baldauf C. and Bauer P. (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine* 24, 3697–3714.

40. Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.

41. Robertson, T., Wright, F., and Dykstra, R. (1988). *Order Restricted Statistical Inference*. New York: Wiley.

42. Silvapulle and Sen (2005). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley-Interscience, Hoboken, N.J.

43. Stylianou, M. and Flournoy, N. (2002). Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics* 58, 171-177.

44. Taves, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clinical Pharmacology Therapeutics* 15, 443-453.

45. Thall, P.F. and Cook, J.D. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 60, 684-693.

46. Thall, P.F., Inoue, L.Y. and Martin, T. G. (2002). Adaptive decision-making in a lymphocyte infusion trial. *Biometrics* 58, 560-568.

47. Thall, P.F., Nguyen, H.Q, and Estey, E.H. (2009). Patient-specific dose finding based on bivariate outcomes and covariates. *Biometrics* 64, 1126-1136.

48. United States Food and Drug Administration. Critical Path Opportunities List. March, 2006. Available at: http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathinitiative/CriticalPathOpportunitiesReports/UCM077258.pdf

49. United States Food and Drug Administration. Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics. Draft Guidance, Feb, 2010. Available at: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm201790.pdf

50. Wetherill, G.B. (1963). Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society B* 25, 1-48.

51. Xiao, C. and Ivanova, A. (2011a). Dose-finding when the target dose is on a plateau of a dose-response curve (submitted to *Statistics in Biopharmaceutical Research*).

52. Xiao, C. and Ivanova, A. (2011b). Adaptive isotonic estimation of the minimum effective and peak doses (submitted to *Journal of Statistical Planning and Inference*).

53. Yuan, Z. and Chappell, R. (2004). Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clinical Trials* 1, 499-508.