A Bayesian Adaptive Design for Multi-Dose, Randomized, Placebo-Controlled Phase I/II Trials

Fang Xie,[1] Yuan Ji,[2] Lothar Tremmel[3]

1: Cephalon, Inc., Frazer, PA, USA

2: The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

3: Incyte, Inc., Wilmington, DE, USA

Corresponding author contact:

Fang Xie, 41 Moores Road, PO Box 4011, Frazer, PA, 19355, USA

Tel: 610-738-6375

Fax: 610-738-6788

Email: fxie@cephalon.com

Acknowledgements

The authors wish to express their gratitude to Drs. John Messina and Jane Tiller for their support and inspiring discussions during the development and implementation of this adaptive design.

Funding

This work was partly supported by Cephalon Inc. Yuan Ji's research was partly supported by NIH grant R01 CA 132897.

Declaration of Conflicting Interests

Fang Xie is an employee of Cephalon Inc. Lothar Tremmel is an former employee of Cephalon Inc.

Abstract

We present a design for a randomized controlled trial (RCT) featuring two simultaneous iterative processes, dose escalation and cohort expansion. Patient enrollment does not need to stop when transitioning from the evaluation of the dose safety and tolerability into the assessment of its efficacy. The cohort expansion used in dose finding is adaptive, based on the interim comparisons between each dose and placebo. A set of Bayesian rules guides the decisions about dose cohort expansion. Operating characteristics of this design have been evaluated by simulations designed to mimic the trial conduct and outcome in a variety of dose toxicity and efficacy scenarios. Simulation studies demonstrated that our proposed adaptive design can reduce the total sample size as compared to the conventional approach. The sample size reduction is more profound in scenarios when the testing doses are not effective. Simulation studies also demonstrated that this proposed adaptive design controls the false positive error rate at the specified level and provides adequate statistical power to detect the treatment effect. Compared to the conventional approach, our proposed adaptive design is better in that it removes ineffective doses, reduces the total sample size, and maintains adequate power for dose finding. The proposed design has been implemented in an ongoing study and software for trial simulation is available at http://odin.mdacc.tmc.edu/~yuanj/soft.html.

Key words: Bayesian, Adaptive design, Dose escalation, Cohort expansion, Dose-finding, Simulation.

1 INTRODUCTION

Various Bayesian adaptive designs have been developed to improve the efficiency of oncology trials [1]-[8]. A large portion of those Bayesian designs originated from the early-stage oncology trials designed to test ascending doses for toxicity in cancer patients. These trials focus on reaching the maximum tolerable dose in the belief that higher doses also yield better responses. The decision rules are either based on a toxicity threshold or putative dose response curves.

Dose-finding methods encompassing both toxicity and efficacy designed for a single-arm trial have been considered by Thall *et al* [2] and further developed in a series of publications and the software "Multc Lean" [9]-[12]. The valuable concept behind this series of work is incorporating both safety and efficacy considerations into the decision rules involved in early-stage clinical development. While this series of work formalizes the concept for dose-finding in practice, its applications are mostly limited to single-arm open-label trials. The existing methods are not readily suitable for dose-finding in the randomized controlled trial (RCT) setting, which requires estimating the difference between the testing treatment and control at different dose levels simultaneously. The RCT is the gold standard that should be used even in early stages of clinical research, whenever possible.

In this paper, we present a Bayesian adaptive design for a particular RCT setting encountered in an ongoing study involving patients suffering from sciatica [13]. This design features two simultaneous iterative processes, dose escalation and cohort expansion. Dose escalation depends on the safety of each dose and is to occur only when the current dose and lower doses are deemed safe. As soon as the first dose is considered safe, a cohort expansion process takes place

to determine enrolling additional patients with a prefixed cohort size based on the observed response rate in the treatment arm and the accumulative placebo data. Although the cohort expansion may still be in progress at a lower dose, the safety testing can advance to a higher dose. The two processes continue simultaneously in order to identify efficacious doses within the safe dose range until the study is stopped because the stopping rules have been met or the maximum cohort size has been reached (see Figure 1 for an illustration).

We do not use binding dose escalation rules because the safety evaluation is a complex process depending on comprehensive medical review; this lack of binding rules differs from the practice followed in some of the existing methods. In reality, safety monitoring is the responsibility of the sponsor and the data and safety monitoring board. In general, one can use model-based dose-escalation schemes and combine them with our proposed adaptive expansion scheme for future designs. For the sake of discussion, we used the "no more than one" rule in dose escalation, i.e. dose escalation is stopped if 2 or more patients have reactions indicating toxicity.

2 STUDY DESIGN

2.1 General Design and Procedure

Let $d = 1, \ldots, D$ represent the doses of a treatment to be tested in an ascending fashion for safety and tolerability. Initially, $n_{1d} + n_{0d}$ patients are randomized, with n_{1d} being treated with the active dose d and n_{0d} treated with the corresponding placebo. If the safety outcome with the n_{1d} patients warrants a dose escalation, a new cohort of $n_{1,d+1} + n_{0,d+1}$ patients is enrolled in the next higher dose level, $d+1$. Meanwhile, the current dose d is evaluated for efficacy, with the possibility of

expanding the cohort size to obtain more efficacy data. The cohort for dose *d* may be expanded according to the Bayesian rules defined by posterior probabilities given the responder data. The cohort expansion rules are applied to all doses deemed safe until an effective dose is identified or the maximum cohort size is reached.

2.2 Probability Model

Let Y_{1d} denote the number of responders among the n_{1d} patients treated with active dose d ($d = 1,\ldots,D$). In conducting an unblinded efficacy analysis, data from placebo-treated patients who have been unblinded already are pooled from all the doses to improve the estimate of the placebo response rate. Let $n_0 = \sum_{d=1}^{D} n_{0d}$ be the total number of patients treated with placebo in all dose cohorts and Y_0 be the number of responders among the n_0 patients. Assume that Y_{1d} ($d = 1,\ldots,D$) and Y_0 are independent random variables following the binomial distributions, $Binom(n_{1d}, p_{1d})$ and $Binom(n_0, p_0)$, respectively. The joint likelihood function for all doses can be expressed as

$$L(p_{11}, \mathbf{K}, p_{1D}, p_0) \propto \prod_{d=1}^{D} p_{1d}^{Y_{1d}} (1 - p_{1d})^{n_{1d} - Y_{1d}} \times p_0^{Y_0} (1 - p_0)^{n_0 - Y_0}.$$

Furthermore, the prior distribution for the responder rates p_{1d} and p_0 is assumed to be $Beta(0.5, 0.5)$, where $Beta(a, b)$ denotes a beta distribution with density $\propto x^{a-1}(1-x)^{b-1}$. With this Beta prior, the posterior distribution of p_{1d} is $Beta(0.5 + Y_{1d}, 0.5 + n_{1d} - Y_{1d})$ and the posterior distribution of p_0 is $Beta(0.5 + Y_0, 0.5 + n_0 - Y_0)$. Note that $Beta(0.5, 0.5)$ is the Jeffreys prior [14] for binomial probabilities and that the Jefferys prior may be more sensitive to rare outcomes

than the uniform prior when sample size is small. The Jeffreys prior suits early-stage trial particularly well since there is no previous clinical experience to inform the assumption about the response rates at this stage.

2.3 Decision Rules

The Bayesian adaptive rules for dose cohort expansion are defined on the basis of the posterior probability

$$R = Prob(p_1 > p_0 + \Delta | data),$$

where $0 \leq \Delta < 1$ denotes the minimum difference between p_1 and p_0 that is needed for decision making. Note that the posterior probability function R does not have a closed form solution in general [15]-[16], although simplifications are possible for a few special cases such as when the prior probability is *Beta(1,1)* and computer software has been developed for numerical calculations of R [17]. With the advanced computing power nowadays, the posterior probability can be estimated well using the direct sampling Monte Carlo method [14].

To construct the dose expansion rules, we first divided the probability space (0, 1) into several mutually exclusive intervals taking the same approach as Ji *et al* [5]. The number of intervals depends on the number of action items one wishes to take. We divide the probability space into three intervals using a pair of boundary parameters, $0 < k_1 < k_2 < 1$, and denote the three intervals as:

$$I_1 = (0, k_1], \quad I_2 = (k_1, k_2), \text{ and } I_3 = [k_2, 1),$$

where k_1 is close to 0 and k_2 is close to 1.

In order to define the Bayesian rules for cohort expansion, several quantities need to be specified at the study design stage. The minimum required treatment response rate (MRT) sets the treatment effect threshold that disqualifies any cohort expansion by reason of futility if the effect is estimated to be below the MRT. The sufficient treatment response rate (STR) sets the satisfactory level of the efficacy signal that requires no further cohort expansion. If the effect exceeds the STR, the study ends because an effective dose has been identified. The maximum tolerable placebo response rate (MTP) sets the placebo response to a level above which is unlikely for the active doses to demonstrate a superior effect.

We propose an iterative cohort expansion procedure during which the following rules are repeatedly applied to the cumulative data for each dose. :

- Rule 1 (placebo response): Calculate the posterior probability $R_1 = Prob(p_0 > MTP|data)$. Halt the study if $R_1 > k_2$ to investigate an excessive placebo response.
- Rule 2 (futility): Calculate the posterior probability $R_{2d} = Prob(p_{1d} < MRT|data)$. Stop the enrollment for dose d if $R_{2d} > k_2$.
- Rule 3 (intermediate): If the stopping criteria are not met by Rule 1 and Rule 2, calculate the posterior probability $R_{3d} = Prob(p_{1d} > p_0 + \Delta|data)$. Assign action items according to the following rules:

$$Decision\ for\ dose\ d = \begin{cases} Stop\ enrollment & if\ R_{3d}\ in\ I_1 \\ Repeat\ a\ cohort & if\ R_{3d}\ in\ I_2 \\ Expand\ to\ maximum\ size & if\ R_{3d}\ in\ I_3. \end{cases}$$

We will denote the three decisions “Stop,” “Repeat,” and “Expand” hereafter.

- Rule 4 (superiority): If the stopping criterion is not met by Rule 1, calculate the posterior probability $R_{4d} = Prob(p_{1d} > STR|data)$. Stop the enrollment for the entire study if $R_{4d} > k_2$.

Of the actions in Rule 3, "Repeat" is a precautionary option to mitigate the risk of expanding the cohort too quickly when the signal is weak from a small initial cohort. One may consolidate "Repeat" with one of the other actions at any step of the iteration. Rule 4 is designed to provide an option to stop the study when a promising efficacy signal is seen and no better efficacy can be expected from other doses. We recommend that Rule 4 be used only when a dose cohort reaches a relatively large size or the maximum to avoid stopping the study prematurely.

In summary, our recommended adaptive design includes the following steps:

1. Apply Rule 1 and Rule 2 before enrolling additional patients in the current dose level for potential early stopping of the dose or the trial.
2. Apply Rule 3 to enroll patients adaptively if the dose is not stopped.
3. When the maximum cohort size has been reached and the response has been observed, apply Rule 4 for potential stopping of the trial because of the potential of a sufficient efficacy signal.

3 THE ONGOING STUDY

The proposed adaptive design has been implemented in an ongoing clinical trial [13]. The study design and operating characteristics of this study design will be presented in this section.

3.1 Study Design

In this study, $D = 5$ dose levels of a new compound are to be examined. Patients will be randomized to an active dose and its corresponding placebo at a 2:1 ratio throughout the study. For a dose *d*, an initial cohort of 9 patients is to be randomized with 6 receiving dose *d* and 3 receiving placebo. Dose escalation will be permitted if no more than one out of 6 patients experiences reactions indicating toxicity. When safety evaluation warrants, we will perform an unblinded responder analysis for dose *d* and apply the three steps of adaptive dose expansion aforementioned. The maximum cohort size to be enrolled is 36 per dose. The unblinded responder analysis will include all patients in the initial cohort for dose *d* and the placebo patients from all previously unblinded cohorts to enhance the estimation of the placebo response for the current analysis. Figure 1 illustrates the simultaneous dose escalation and cohort expansion processes in this study.

A slight modification has been made to reduce logistical burden involved in conducting the trial: When the decision to "Repeat" has already been made regarding a dose, in the next expansion for that dose, only two possible decisions will be considered, "Stop" or "Expand." That is, in the second stage of dose expansion, we will either expand to the maximum number of patients for that dose or stop the dose. This modification is intended purely to reduce the logistical load in practice and does not need to be implemented in other trials.

The study parameters described in Section 2.3 are as follows: the MTP will be set at 0.5, the MRT at 0.2, the STR at 0.8, the targeted difference in treatment response rate (Δ) at 0.2, and the two boundary parameters associated with the decision rules at $k_1 = 0.1$ and $k_2 = 0.9$.

3.2 An Illustration

In this section, we present a set of hypothetical trial outcomes of an artificial trial to illustrate the how the cohort expansion rules are applied. In this artificial setting, all 5 doses were assumed to be safe for dose escalation and the true response rates are arbitrarily chosen. We performed a total of 10 interim responder analyses based on randomly generated responses and applied the cohort expansion rules in this hypothetical trial. The decisions and the trial process are summarized in Table 1. Along with the observed outcomes Y's for each dose d, the four posterior probabilities R's as defined in Section 2 are displayed under "Outcome" in Table 1. To recapitulate, R_1 was the posterior probability of an excessively high placebo response rate leading to halting of the trial, R_{2d} was the posterior probability of stopping individual dose for futility, R_{3d} was the posterior probability for individual dose cohort expansion, and R_{4d} was the posterior probability of declaring efficacy and stopping the entire trial. In this table, the last column shows the adaptive decisions made based on the observed outcomes.

Following the results presented in Table 1, this artificial trial ended with a total enrollment of 126 patients to conclude the dose finding. Dose 3 (d_3) was identified as an effective dose, with 22 responders out of 24 patients treated. After analysis #10 for d_3, the study enrollment was stopped since d_3 satisfied Rule 4 (superiority), with $R_{4d} = 0.934$. Table 1 also shows that both

doses d_1 and d_2 were effective, given that R_{3d} was 0.823 in analysis #4 for d_1 and 0.935 in analysis #8 for d_2. Rule 4 was applied to doses d_1 and d_2 since their cohort size had reached the maximum. Although doses d_1 and d_2 also were expanded to the maximum size and had high response rates in analyses #4 and #8, they did not stop the study because for neither dose was the posterior probability R_{4d} greater than the STR level (0.8) under Rule 4. Lastly, in analysis #9, dose d_5 was stopped for futility by Rule 2 since $R_{2d} = 0.905 > k_2$.

4 SIMULATION STUDIES

4.1 Simulation Settings

We assessed the operating characteristics of the proposed design using the ongoing study through computer simulations. In all simulations, the design parameters were set to be constant: MTP=0.5, MRT=0.2, STR=0.8, Δ=0.2, k_1=0.1 and k_2=0.9. We considered two sets of toxicity situations: equal toxicity rates, with $t_d = 0.05$ for all $d=1,...,5$ doses, and increasing toxicity rates (t_d's) of (0.03, 0.06, 0.09, 0.12, 0.15) for doses 1 through 5, respectively. Two placebo response rates were used in the simulations: 0.2 and 0.5. In Table 2, we constructed 12 scenarios with different combinations of the true treatment response rates (p_{1d}) and placebo response rates (p_0). For each of the 24 combinations of toxicity and responses, 10,000 simulation trials were conducted using the proposed adaptive design.

In the simulation, 9 initial samples were drawn, 6 from $Binom(6, p_{1d})$ and 3 from $Binom(3, p_0)$. Responses indicating toxicity were generated from $Bionom(6, t_d)$. If the toxicity incidence was ≤ 1 among the 6 treated patients for dose d, we proceeded with the unblinded responder analysis

and applied the rules specified in Section 2.3. Available unblinded placebo data were pooled for each analysis. If the toxicity incidence was >1, enrollment to the subsequent higher dose levels was suspended. Of the posterior probabilities needed to apply the rules specified in Section 2.3, R_1, R_{2d}, and R_{4d} were directly calculated. The posterior probability $R_{3d} = Prob(p_{1d} > p_0 + 0.2|data)$ was estimated by averaging 1,000 repeated random draws of p_{1d} and p_0 from their posterior distributions.

For each of the scenarios in Table 2, we compared the proposed adaptive design to a conventional approach that consisted of a dose-escalation safety study using a 3+3 design and a subsequent dose-finding study using a parallel-group design. The parallel-group design is a popular scheme for studying the treatment effect with multiple dose levels within a safe range. In a recently published study utilizing the parallel-group design [18], more than 400 patients were randomized to a total of 9 groups (at over 40 per group) including 1 placebo group, 3 doses within regimen 1, 4 within regimen 2, plus 1 active control group. In our simulation work, for comparison purposes, the total sample size for the conventional approach was set to match the maximum sample size using the proposed adaptive design. The maximum sample size using our proposed adaptive design was 180 if the cohort size reached 36 patients for all 5 dose levels. Considering placebo as dose 0, there were 6 groups in total in the conventional approach, and thus the sample size was 30 patients (=180/6) per group. Since some degree of toxicity was assumed in the simulation (Section 4.1), it was possible for dose escalation to end due to the toxicity criterion in a small number of simulated trials. Therefore, the average sample size for the conventional approach was expected to be slightly lower than but close to 180.

4.2 Simulation Results

The mean total sample size using the conventional design was 176 or 168 when toxicity was equal, with $t_d = 0.05$, or increasing, with $t_d = (0.03, 0.06, 0.09, 0.12, 0.15)$, for d=1,…,5, respectively. Compared to the conventional approach, the mean total sample size using the adaptive design was considerably lower in all 24 simulation scenarios (Table 3), with the sample size reduction ranging from approximately 15%-50% in most cases. Specifically, the adaptive design was able to save over 40% on sample size in all null scenarios. If all doses were equally effective, the difference between the two design strategies decreased to 12%-15% when $p_0 = 0.2$. Saving on sample size was more substantial in the case of a high placebo response rate ($p_0 = 0.5$) than a low placebo response rate ($p_0 = 0.2$) because the placebo response stopping rule and the futility stopping rule were more likely to be invoked in the high situations.

In Figure 2 and Figure 3, we present the mean cohort size per dose, i.e., the average number of patients treated at each dose. The mean cohort size per dose was in accordance with the dose-response pattern and toxicity pattern assumed in the simulations. The cohort size was approximately 18-20 when all doses and placebo were equal ("Null" scenarios), indicating that the adaptive design was able to stop the cohort expansion in the event of ineffective doses right after repeating the initial cohort size in most scenarios. In contrast to this substantial saving, the conventional design enrolled nearly 180 patients regardless. The adaptive design provided the least sample size reduction when all doses were marginally better than the required minimum difference in response rate ($\Delta = 0.2$) for cohort expansion ("Equal" scenarios).

The superiority rule (Rule 4) was applied in the simulations when a dose cohort reached the maximum size of 36. Simulation results, summarized in Table 4, showed a decent probability of stopping the study using the adaptive design when a sufficiently high rate of response to the treatment ($p_{1d} = 0.9$) was assumed for at least one dose in increasing, decreasing, and u-shape dose response settings when $p_0 = 0.5$.

4.3 Comparison to Conventional Design

We carefully assessed and compared the operating characteristics of the proposed adaptive design to those of the conventional approach. To ensure a fair comparison, we took an approach similar to that of comparing two hypothesis testing procedures, in which one would first match the type I error rate of the two tests under the null hypothesis, and compare the power of the tests under various alternative hypotheses. In our case, we designed a null scenario and calibrated the adaptive design so that the two designs give the same *trial type I error rate*, a quantity to be defined below. We then examine the trial power (again defined next) for various alternative scenarios.

As demonstrated earlier, the proposed adaptive design in general requires smaller sample sizes. It is desirable to reduce the sample size at an affordable expense of power loss for dose finding. For dose selection, the simple hypothesis H_0: $p_{1d} = p_0$ versus H_1: $p_{1d} > p_0$ was tested for individual doses separately. The trial Type I error was defined as the probability of rejecting H_0 in favor of H_1 incorrectly for at least one dose under the Null scenario setting in Table 2. Correspondingly, the trial power was defined as the probability of correctly rejecting H_0 in favor of H_1 for at least one dose in the trial. In order to compare the trial power of the two different methods, the trial

Type I error needed to be set at a common level for the null scenario, in which all the doses had the same efficacy rate as the placebo. We first set the Type I error for individual doses as α = 0.05 using the 2-sided Chi-square test in the conventional approach. We rejected H_0 if the lower limit of the $(1-\alpha)100\%$ confidence interval for $p_{1d} - p_0$ was above 0. The mean trial Type I error rate for the conventional approach was then calculated from the 10,000 simulations in the null scenario setting as the proportion of the simulations in which the H_0 was rejected for at least one dose. Next, we calibrated a cutoff value k_α^* using the Bayesian test for individual doses in the null scenario setting, with which we rejected H_0 if the posterior probability $Prob(p_1 > p_0|data) \geq (1 - k_\alpha^*/2)$. The calibration could be achieved by performing the Bayesian test on the 10,000 simulated trials with various k_α^* values iteratively until the mean trial Type I error *rate* matched that for the conventional approach using the Chi-square test. Note that k_α^* was a function of p_0. From the 10,000 simulations using the ongoing study design, k_α^* was found to be 0.072 and 0.064 corresponding to the $p_0 = 0.2$ and $p_0 = 0.5$ settings. The simulated power results for $p_0 = 0.2$ are summarized in Table 5 for flat toxicity and in Table 6 for increasing toxicity. It can be seen that under the two calibrated k_α^* values, the two designs had the same trial Type I error rate for the null scenario.

When $p_{1d} = p_0$ (null scenarios), the mean trial Type I error was approximately 0.133. In all other scenarios, $p_{1d} > p_0$ was true for at least one dose. The mean trial power of the adaptive design was generally high (>0.85); however, it was 5%-10% lower than that of the conventional approach. As indicated by Table 3, this lowered trial power was the tradeoff for a 25% savings on the sample size using the adaptive design. The mean power for individual doses using the

adaptive design was generally similar to or less than 0.1 lower than that of the conventional design. A greater difference in power (>0.1) was observed for individual doses when the difference between p_{1d} and p_0 reached 0.4, which occurs in the n-shape and u-shape scenarios.

5 DISCUSSION AND CONCLUSION

We propose a Bayesian adaptive design to evaluate safety and efficacy simultaneously in a placebo-controlled RCT setting to enhance the efficiency of the early-stage clinical development. This adaptive design introduces several parameters that act as gatekeepers in the decision-making process when little is known about the testing compound and sample size is limited in the early-stage trials. The choice of the maximum tolerable placebo response rate (MTP) depends on the disease to be treated and thus requires clinical expertise. It could be set relatively high depending on the disease and indication of the new compound under testing. One example is the well known high rate of response to placebo in the psychiatric therapeutic area. Another example is the response to the experimental compound to be used as an add-on therapy. In this case, the placebo effect indeed represents the efficacy of the standard of care or the background therapies.

Of the two boundary parameters k_1 and k_2, parameter k_2 is more important for decision-making and should be chosen on the basis of thorough simulation studies. In principle, k_2 should be set close to 1. However, one needs to be mindful about balancing the vigor of the statistical inference and the usefulness of the adaptive rules. For an early-stage trial with a relatively small sample size, k_2 could be set too high to provoke any decision rules, compromising the advantages of the adaptive design. As an example, let us consider the illustrative trial in section 3.2 and

raise k_2 to 0.95 for the same trial outcomes in Table 1. From the interim analysis #5 of dose 3 (d_3), $R_{3d} = Prob(p_{1d} > p_0 + 0.2|data) = 0.946 < 0.95$. With $k_2 = 0.95$, the decision after this analysis would be changed to "Repeat" d_3 with 9 patients, as opposed to "Expand" d_3 to the maximum with 27 new patients. Moreover, the result of analysis #10 of d_3 with the maximum cohort size of 36 would fail to stop the study by Rule 4 for reaching the sufficient treatment response (STR) because $R_{4d} = Prob(p_{1d} > STR|data) = 0.934 < 0.95$. Instead, the trial would go on, with two more rounds of cohort expansion for dose d_4 to the maximum size of 36, before the reaching final conclusion that both doses d_4 and d_5 are ineffective. In this case, raising k_2 to 0.95 would increase cost and time so that 27 more patients could be added but result in the same conclusion as that under $k_2 = 0.90$.

Using the ongoing study as an example, the proposed adaptive design will produce a sizeable cohort of placebo patients that may be adequate to provide a reliable estimate of the placebo response rate. One may argue about the need to randomize more patients to receive placebo treatment after reaching a set enrollment target. Two practical considerations support the randomization throughout the study. The rate of response to placebo may rise with the dose escalation if it is impossible to blind the dose level for dose-escalation purposes. The placebo effect also could vary in multi-center studies, as is the case with the ongoing study. Therefore, it is a good practice to continue enrolling placebo patients at every step of the cohort expansion in studies conducted in a multi-center setting or with difficulties to blind the dose levels. Regardless of the response to placebo, the pooling of placebo patients can ensure the placebo response rate is estimated from all centers for all steps of dose escalation and cohort expansion.

Careful planning and executable logistics for the adaptive design will be critical to benefit fully from the proposed design. The implementation of the proposed adaptive design will be more complex than the conventional, parallel-group dose-finding studies. The logistics is even more challenging in a multi-center setting, as is the case with the ongoing study with the new adaptive design. The study will involve multiple iterations of interim analysis and decision making. It is possible to have multiple doses at different stages of cohort expansion simultaneously. Proceeding with dose-cohort expansion sequentially has apparent advantages. This approach is consistent with the desire to identify a minimum effective dose and can simplify the patient allocation logistics.

Our proposed adaptive design can be modified to suit a variety of situations. Future designs can incorporate more advanced features such as adaptive randomization or simultaneous cohort expansion for all safe doses. Our experience indicates that it is important to balance between the advantages of adaptive trial designs and the increased complexity in trial conduct and logistics. We consider the particular design for the ongoing study as a good compromise between the advanced statistical modeling and constraint in practice.

In conclusion, the proposed adaptive design can accelerate the clinical development program and lower the number of patients needed to study. Simulation studies demonstrated that the proposed adaptive design has sound statistical properties in controlling the false positive error rate at the specified level and providing adequate statistical power to detect the treatment effect. Compared to the conventional approach, this adaptive design is expected to be more efficient in that it

removes ineffective doses, allocates more patients to the effective doses, and reduces the overall sample size.

6 REFERENCES

[1] Berry SM, Carlin BP, Lee JJ, Muller P. Bayesian Adaptive Methods for Clinical Trials. CRC press, 2010.

[2] Thall PF and Simon RM. Practical Bayesian Guidelines for Phase IIB clinical Trials. *Biometrics* 1994; 50: 337-349

[3] O'Quigley J, Pepe M and Fisher L. Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; 46: 33-48.

[4] Ji Y, Li Y and Bekele BN. Dose-finding in phase I oncology clinical trials based on toxicity probability intervals. *Clinical Trials* 2007; 4: 235-244.

[5] Ji Y, Liu P, Li Y, Nebiyou BN. A modified toxicity probability interval method for dose-finding trials. *Clinical Trials* 2010; 7: 653-663.

[6] Hu B, Bekele BN, Ji Y. Adaptive Dose Insertion in Early Phase Clinical Trials. *Clinical Trials* 2010; Sep. 6, Epub; In Press.

[7] Ji Y, Feng L, Shpall EJ, Kebriaei P, Champlin R, Berry D, Cooper LJN Bayesian Continual Reassessment Method for Dose-Finding Trials Infusing T Cells with Limited Sample Size. *Journal of Biopharmaceutical Statistics* 2011, in press.

[8] Ji Y, Bekele BN (2009), Adaptive randomization for multi-arm comparative clinical trials based on joint efficacy/toxicity outcomes, *Biometrics* 65(3):876-84

[9] Thall PF, Simon R, and Estey EH. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statistics in Medicine* 1995; 14:357-379.

[10] Thall PF and Sung HG. Some extensions and applications of Bayesian strategy for monitoring multiple outcomes in clinical trials. *Statistics in Medicine* 1998; 17:1563-1580.

[11] Thall PF and Cook JD. Adaptive dose selection using efficacy-toxicity trade-offs: illustrations and practical considerations. *Journal of Biopharm Statistics* 2006; 16: 623-638.

[12] Cook JD. Multc Lean Statistical Tutorial. 2005.

[13] A Randomized, Double-Blind, Placebo-Controlled, Ascending-Dose Study to Evaluate the Safety and Efficacy of CEP-37247 Administered by the Transforaminal Epidural Route for the Treatment of Patients With Lumbosacral Radicular Pain Associated With Disk Herniation. *www.Clinicaltrials.gov* identifier: NCT01240876. Sponsored by Cephalon, Inc.

[14] Carlin BP and Louis TA. Bayesian Methods for data analysis, 3rd Edition, CRC Press 2009

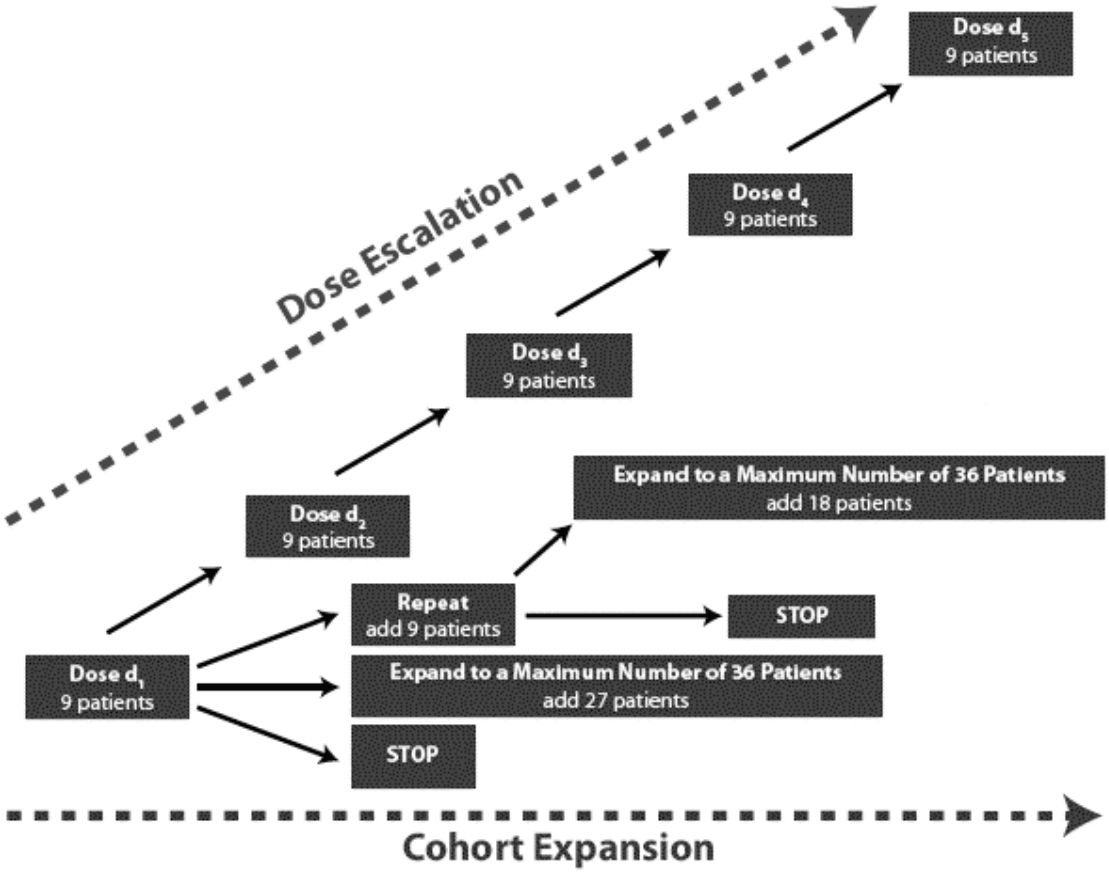
[15] Nadarajah S. Reliability for some bivariate distributions. *Mathematical Problems in Engineering* 2005: 101–111

[16] Cook JD and Nadarajah S. Stochastic inequality probabilities for adaptively randomized clinical trials. *Biometrical Journal* 2006; 48(3): 356-365.

[17] Wathen JK, Nguyen HQ, and Cook JD. Inequality calculator. 2005. http://biostatistics.mdanderson.org/softwaredownload.

[18] McClung MR, Lewiecki EM, Cohen SB, *et al*, Denosumab in Postmenopausal Women with Low Bone Mineral Density, *New England Journal of Medicine* 2006; 354(8), 821-831.

Figure 1: Dose Expansion and Escalation Schema



Note: patients will be randomized to the active dose and its corresponding placebo arm at a 2:1 ratio for each dose.

Figure 2: Mean Cohort Size Assuming Equal Toxicity (0.05, 0.05, 0.05, 0.05, 0.05)

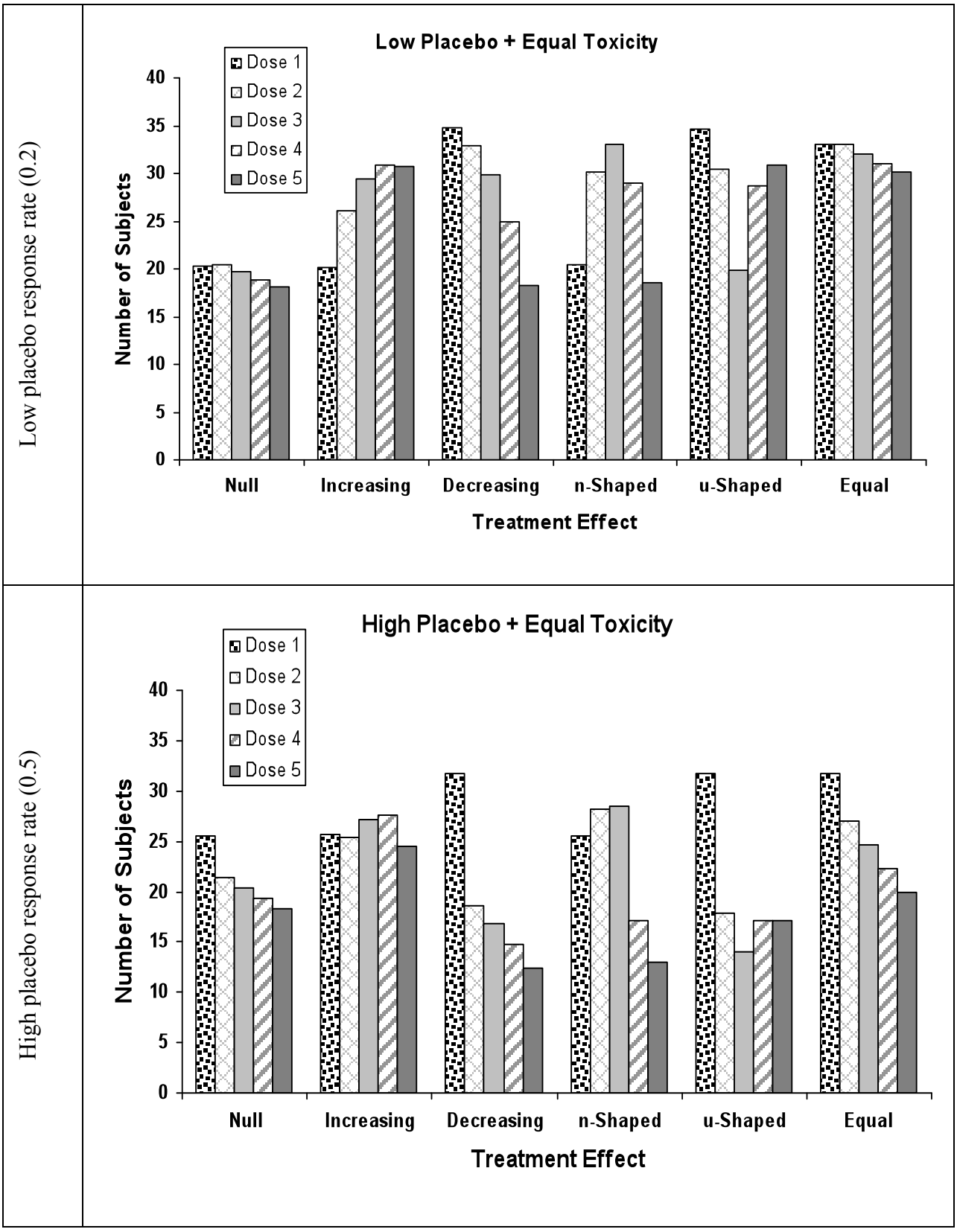


Figure 3: Mean Cohort Size Assuming Increasing Toxicity (0.03, 0.06, 0.09, 0.12, 0.15)

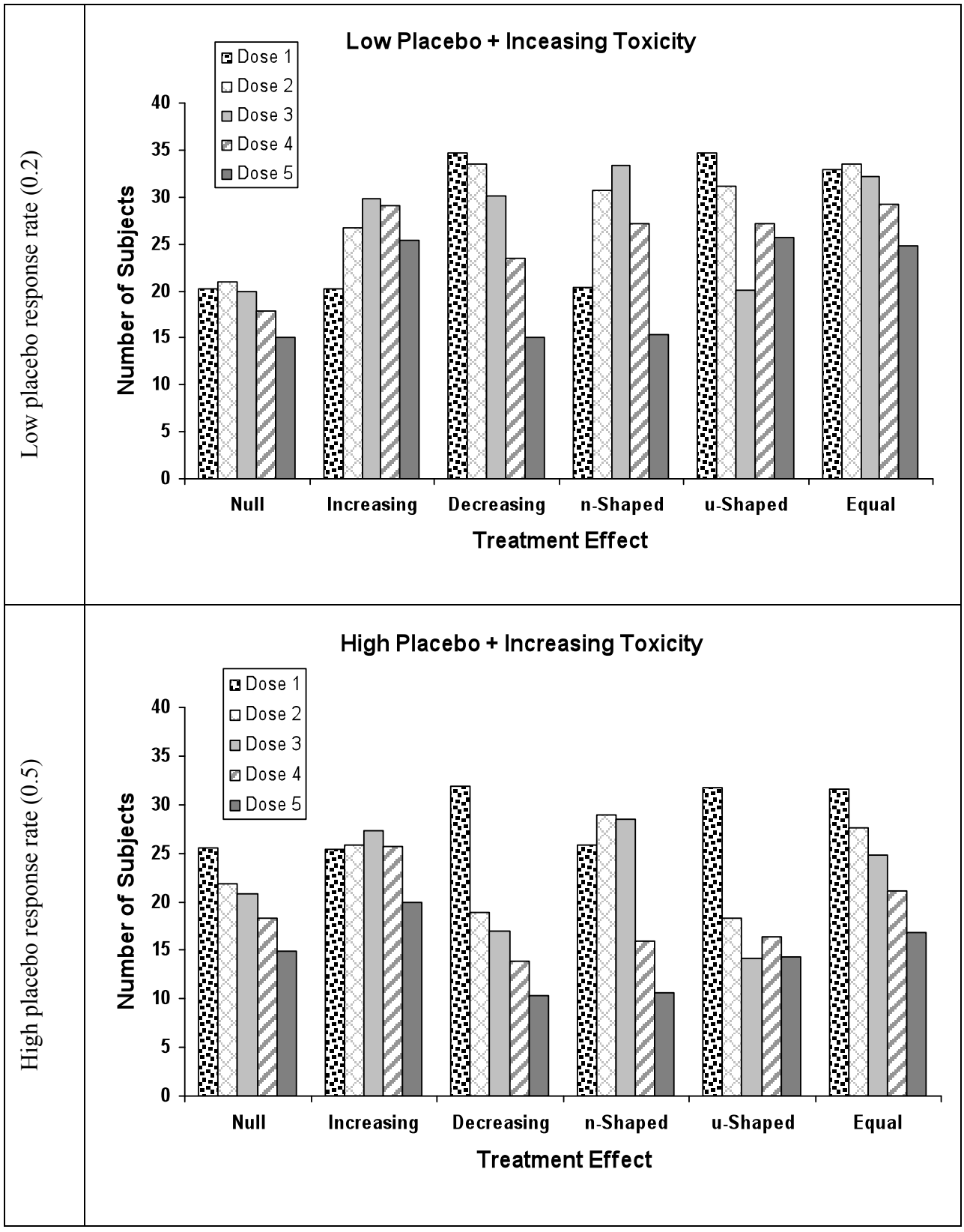


Table 1: Responder Analyses and Decisions Made in an Illustrative Trial

Interim Analysis		Placebo Data*		Dose Data		Outcome				
#	***Dose (d)***	Y_0	N_0	Y_{1d}	N_{1d}	R_1	R_{2d}	R_{3d}	R_{4d}	Decision
1	1	0	3	2	6	0.033	0.196	0.563	.	Repeat
2	1	2	9	6	12	0.045	0.009	0.612	.	Expand
3	2	2	9	3	6	0.045	0.044	0.603	.	Repeat
4	1	5	18	15	24	0.028	0.000	**0.823**	0.022	End Dose
5	3	5	18	5	6	0.028	0.000	0.946	.	Expand
6	2	7	24	8	12	0.019	0.000	0.842	.	Expand
7	4	7	24	2	6	0.019	0.196	0.221	.	Repeat
8	2	9	33	16	24	0.004	0.000	**0.935**	0.058	End Dose
9	5	9	33	0	6	0.004	**0.905**	0.007	.	End Dose
10	3	12	42	**22**	**24**	0.002	0.000	1.000	**0.934**	Stop Study

*Pooled placebo data across all doses at the time of analysis.

Table 2: Treatment Response Rates (p_{1d}) in the Simulations

Treatment Response Rate (p_{1d}) Pattern (d_1,d_2,d_3,d_4,d_5)	Placebo Response Rate (p_0) 0.2	0.5
Null	0.2, 0.2, 0.2, 0.2, 0.2	0.5, 0.5, 0.5, 0.5, 0.5
Increasing	0.2, 0.3, 0.4, 0.5, 0.6	0.5, 0.6, 0.7, 0.8, 0.9
Decreasing	0.6, 0.5, 0.4, 0.3, 0.2	0.9, 0.8, 0.7, 0.6, 0.5
n-shape	0.2, 0.4, 0.6, 0.4, 0.2	0.5, 0.7, 0.9, 0.7, 0.5
u-shape	0.6, 0.4, 0.2, 0.4, 0.6	0.9, 0.7, 0.5, 0.7, 0.9
Equal	0.5, 0.5, 0.5, 0.5, 0.5	0.8, 0.8, 0.8, 0.8, 0.8

Table 3: Mean Total Sample Size and Percentage Reduction (in parentheses) Using the Adaptive Design Compared to the Size Needed Using the Conventional Approach

	Equal Toxicity[1] (0.05, 0.05, 0.05, 0.05, 0.05)						Increasing Toxicity[2] (0.03, 0.06, 0.09, 0.12, 0.15)					
Placebo Response Rate	Null	Increasing	Decreasing	n-Shape	u-Shape	Equal	Null	Increasing	Decreasing	n-Shape	u-Shape	Equal
0.2	97	137	141	131	145	159	94	131	137	127	139	153
	(46)	(24)	(22)	(27)	(20)	(12)	(48)	(27)	(24)	(29)	(23)	(15)
0.5	105	130	94	112	98	126	101	124	92	110	95	122
	(42)	(28)	(48)	(38)	(46)	(30)	(44)	(31)	(49)	(39)	(47)	(32)

[1]Compared to 176 using the conventional approach.
[2]Compared to 168 using the conventional approach.

Table 4: Stopping probabilities induced by Rule 4, $R_{4d} = Prob(p_{1d} > 0.8|data) \geq 0.9$, When Maximum Cohort Size Is Reached

p_1 (d_1,d_2,d_3,d_4,d_5)	$p_0 = 0.2$					$p_0 = 0.5$				
	d_1	d_2	d_3	d_4	d_5	d_1	d_2	d_3	d_4	d_5
Null	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
Increasing	0.00	0.00	0.00	0.01	0.05	0.01	0.05	0.16	0.36	**0.50***
Decreasing	0.06	0.01	0.00	0.00	0.00	**0.69***	0.19	0.07	0.02	0.00
n-shape	0.00	0.00	0.06	0.00	0.00	0.01	0.16	0.61	0.07	0.00
u-shape	0.06	0.00	0.00	0.00	0.05	**0.70***	0.07	0.00	0.07	**0.28***
Equal	0.01	0.01	0.01	0.01	0.01	0.41	0.34	0.30	0.26	0.22

*$p_{1d} = 0.9$ in simulation settings.

Table 5: Simulated Type I Error/Power Rate for $p_0 = 0.2$ ($k_\alpha^* = 0.072$) and Equal Toxicity

p_1 (d_1,d_2,d_3,d_4,d_5)	Design	Dose					Trial
		d_1	d_2	d_3	d_4	d_5	
Null	Adaptive	0.0394	0.0345	0.0334	0.0301	0.0299	0.1328
	Conventional	0.0292	0.0299	0.0292	0.0263	0.0299	0.1328
Increasing	Adaptive	0.0344	0.1293	0.3658	0.6473	0.8165	0.8874
	Conventional	0.0274	0.1644	0.4257	0.7151	0.9072	0.9853
Decreasing	Adaptive	0.7193	0.6395	0.4115	0.1697	0.0314	0.8761
	Conventional	0.9066	0.7129	0.4307	0.1591	0.0349	0.9860
n-shape	Adaptive	0.0376	0.3323	0.8152	0.4051	0.0323	0.8578
	Conventional	0.0285	0.4133	0.9080	0.4270	0.0321	0.9647
u-shape	Adaptive	0.7230	0.3706	0.0338	0.4040	0.8108	0.9455
	Conventional	0.9060	0.4154	0.0289	0.4160	0.9070	0.9955
Equal	Adaptive	0.4997	0.6350	0.6726	0.6871	0.6869	0.9450
	Conventional	0.7052	0.7073	0.7066	0.7064	0.7002	0.9953

Table 6: Simulated Type I Error/Power Rate for $p_0 = 0.2$ ($k_\alpha^* = 0.072$) and Increasing Toxicity

p_1 (d_1,d_2,d_3,d_4,d_5)	Design	Dose					Trial
		d_1	d_2	d_3	d_4	d_5	
Null	Adaptive	0.0397	0.0365	0.0316	0.0303	0.0259	0.1323
	Conventional	0.0320	0.0295	0.0296	0.0317	0.0320	0.1336
Increasing	Adaptive	0.0396	0.1345	0.3695	0.6230	0.6757	0.8467
	Conventional	0.0293	0.1586	0.4040	0.6583	0.8370	0.9087
Decreasing	Adaptive	0.7173	0.6431	0.4201	0.1632	0.0241	0.8784
	Conventional	0.9094	0.7194	0.4389	0.2103	0.0929	0.9858
n-shape	Adaptive	0.0376	0.3334	0.8174	0.3771	0.0253	0.8617
	Conventional	0.0307	0.4132	0.8691	0.4406	0.0940	0.9236
u-shape	Adaptive	0.7278	0.3918	0.0351	0.3955	0.6685	0.9294
	Conventional	0.9075	0.4325	0.0668	0.4209	0.8390	0.9922
Equal	Adaptive	0.4993	0.6348	0.6813	0.6436	0.5598	0.9483
	Conventional	0.7178	0.7138	0.7115	0.7112	0.7176	0.9876