Published in final edited form as:

J Biopharm Stat. 2013 ; 23(5): 1124–1154. doi:10.1080/10543406.2013.813519.

An adaptive two-stage dose-response design method for establishing Proof of Concept

Yoko Franchetti*, Stewart J. Anderson[†], and Allan R. Sampson[‡]

Yoko Franchetti: yoko@jimmy.harvard.edu

*Dana-Faber Cancer Institute and Harvard School of Public Health Boston, MA, 02215, USA

[†]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA

[‡]Department of Statistics, School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA 15261, USA

Abstract

We propose an adaptive two-stage dose-response design where a pre-specified adaptation rule is used to add and/or drop treatment arms between the stages. We extend the multiple comparison procedures-modeling (MCP-Mod) approach into a two-stage design. In each stage, we use the same set of candidate dose-response models and test for a dose-response relationship or proof of concept (PoC) via model-associated statistics. The stage-wise test results are then combined to establish ‘global’ PoC using a conditional error function. Our simulation studies showed good and more robust power in our design method compared to conventional and fixed designs.

Keywords

two-stage adaptive design; adding and/or dropping treatment arms; xMCP-Mod; Proof of Concept; drug development

1 Introduction

In recent years, novel designs of clinical trials have played an increasingly important role in the drug approval process. For example, the use of *adaptive* designs generally allows for earlier decisions whether to continue or stop a trial which combines phases such as Phases I and II or Phases II and III. The FDA (2010) released a draft guidance on adaptive designs and commented that employing adaptive designs in clinical trials is expected to lead to operational efficiency, as well as more informative results given a well-controlled study setting. In the long run, it is expected to benefit patients who need efficacious and safe drugs as early as possible.

The term, ‘adaptive design’, has a broad meaning. Gallo et al. (2006) referred to an adaptive design as being a clinical study design that uses accumulating data to decide how to modify aspects of the study as it continues without undermining the validity and integrity of the

trial. Types of adaptations include adjusting sample size, stopping early due to efficacy or futility, and changing the timing and the number of analyses (Chang et al., 2007). In the FDA's 2010 draft guidance, clinical researchers are encouraged to detail *pre-specified* adaptation rules in the protocol.

In clinical drug development, one of the important factors to consider is the nature of a dose-response (D-R) relationship, as a D-R relationship is recognized as important to determine a maximum tolerated dose, a minimum efficacious dose, and a therapeutic dose window of a drug in its clinical stage. The International Conference on Harmonization Guideline E4 (ICH-E4) on D-R information encourages regulatory agencies and drug developers to employ new approaches in the design and analysis of D-R data for precise determination of key clinical dosages. Accordingly, *adaptive D-R designs* have been one such novel approach of recent use in the pharmaceutical industry.

In certain situations, it is necessary to pre-specify doses to use in clinical trials for D-R research. The reasons include regulation requirements for investigational drugs called Good Manufacturing Practice and corporate quality assurance (QA) policies for investigational drugs. For example, in Japan investigational drugs have been required to be manufactured with the same or very similar specifications to the future drugs' market specifications (Hiyama, 2007). As consistency and equivalency between an investigational drug and its marketed version should be assured, there is hesitation to study adding a new investigational dose that was never used in a previous clinical trial. On the other hand, the ICH-E4 encourages employing new approaches for D-R research, and adding and/or dropping treatment doses in adaptive designs is one such approach.

Contribution—In this paper, we propose a novel adaptive two-stage design which features a framework for pre-specified adaptation between the two stages using a well-characterized set of possible rules. We focus on establishing proof of evidence, or *proof of concept* (PoC), the first step in dose-finding research. Our method aims at establishing *global* PoC by combining the results from both stages while enabling adding and/or dropping doses within one trial given an adaptation rule. We call our method xMCP-Mod as it extends the multiple comparison procedures-modeling (MCP-Mod) method (Bretz et al., 2005 and Pinheiro et al., 2006), and we call an adaptation rule in our framework 'ADTAR' which stands for 'adding and dropping treatment adaptation rule'.

Our xMCP-Mod method essentially consists of two stages of the original MCP-Mod method, and adds adaptive analysis as well as the dose adaptation between the two stages. We set up an adaptive null hypothesis for Stage 2 to resolve sources of dependency of Stage 2 on Stage 1. Remaining data-dependent sample size imbalances due to adaptation are addressed through a weighting procedure. Importantly, our approach controls the family-wise error rate (FWER). The xMCP-Mod method requires the pre-specification of all potential doses and the dose adaptation procedure prior to the start of the trial. Thus, it simplifies complying with regional (non-U.S.) chemistry and manufacturing requirements for investigational drugs or possibly stricter corporate QA policies. Furthermore, it is adhering to the FDA's draft guidance recommending pre-specification.

In this paper we limit the discussion to establishing PoC. The underlying assumption of our study is that in early exploratory trials it may be necessary or preferable to first establish PoC before expending resources on estimating the optimal dose and trying out many doses. The evaluation of adaptive designs by the Pharmaceutical Innovation Steering Committee (PISC) of the Pharmaceutical Research and Manufacturers of America (PhRMA) (Bornkamp et al., 2007, Pinheiro et al., 2010) concluded that establishing PoC is an easier task than estimating a true D-R curve and estimating a target dose. However, it is a necessary step that warrants attention. From the method stand point, an adaptive design that is tailored for establishing PoC given our dose adaptation rule requires a different methodology than an adaptive design optimized for estimating the best dose-response curve or an optimal dose. The data collected in this process may be used to perform dose estimation with other methods once the PoC is established. We follow a commonly used assumption for adaptive designs that the endpoint is a short-term endpoint. Since a PoC study is generally used for an early phase of clinical development, we further assume a small sample size of patients with a non-rare disease so that patient accrual is not an issue. In addition, our method works under the assumption that a dose that is to be used due to a dose adaptation rule is safe, which leads to not allowing a deviation from the rule. Note that this does not imply that all doses are assumed safe from the beginning of the trial but only that they are safe when selected.

We perform extensive simulations and show that our xMCP-Mod method for a given adaptation rule is robust and has good power. In most cases it has the second-best power across all evaluated designs while the other designs are power sensitive to the true D-R relationship. For instance, the original MCP-Mod method can have either the best or the worst power depending on the specified underlying D-R curve. We further establish that our proposed weighting procedure compensates for sample size imbalances due to adaptation and improves power.

Related work—We discuss the most closely related work on adding and/or dropping treatment arms in adaptive designs and extensions of the MCP-Mod method. A general overview of adaptive designs and the MCP-Mod method is provided in Sections 2.2 and 2.3, respectively. The general idea of adding and dropping treatment arms in designs has been discussed and multiple authors have suggested its advantage. However, to our knowledge a framework for adding and/or dropping doses *while at the same time pre-specifying the doses* has not been extensively dealt with in the literature. Enabling pre-specification and dose adaptivity at the same time for establishing PoC is at the core of our paper.

Phillips (2006) commented that the flexibility in adding treatment arms or dropping treatment arms based on efficacy or safety information is generally acceptable and considered to be advantageous. The biggest concerns are how to control the overall type I error and how to appropriately estimate confidence intervals. Additionally, in a D-R study, how to integrate D-R modeling in an adaptive design is another issue.

Both Hung et al. (2006) and Hommel (2001) discussed the modification of dropping arm(s) and overall type I error in terms of testing an effective treatment dose. Hung et al. commented that it may be advisable to redistribute the unused sample size of the dropped

arms to the rest of the retained arms. This may allow one to increase statistical power in detecting an effective dose. Bornkamp et al. (2007) and Pinheiro et al. (2010) studied the power of new D-R study designs with respect to the general objectives of D-R research including the MCP-Mod method and its extensions.

There are three adaptive designs for the objectives of D-R research that are motivated from or extending the MCP-Mod approach. The method proposed by Miller (2010) was the most similar to our proposed xMCP-Mod method. It was an adaptive two-stage design that employed the MCP-Mod in each stage. To detect an upward trend the example adaptation rule picks three doses out of five candidate doses for stage two (to be used together with the placebo), based on the estimated effect sizes among doses of the first stage. They are not utilizing a set of candidate D-R model responses in the adaptation rule. To control the FWER, an adaptive multiple contrast test was used with a conditional error function and an inverse normal combination test. Different contrasts were used in the two stages and a critical value is derived from the doses used in both stages. Their example rule keeps the sample size constant across stages and doses, avoiding issues due to sample size imbalance. The authors did not discuss the application in the more general setting we treat. In contrast to Miller (2010), our proposed adaptation rules are pre-specified and deviations are not possible.

A different extension of the MCP-Mod method to multistage designs without formally controlling the FWER was investigated by Dragalin et al. (2010), using optimal design theory and Bayesian update rules. Instead of using multiple candidate D-R models, Dragalin et al. used a single flexible D-R model for optimizing the design with interim data in order to update the assumed D-R relationship (D-optimization) as well as reallocate patients to different doses. C-optimization was applied for the rest of the study data in order to estimate the minimum efficacious dose. As another extension of the MCP-Mod called aMCP-Mod, Bornkamp et al. (2011) employed the Bayesian theory in the MCP-Mod method to update the D-R parameters based on interim data. New doses for subsequent stages are determined based on the upper and lower boundaries of the updated parameters and thus the optimal dose was determined. Dragalin et al. (2010) commented that with this approach one can find an compromise design, which is robust to model misspecification. Since we do not use a Bayesian approach, our approach is different from these methods.

Synopsis—In Sections 2.1, 2.2, and 2.3, we review general D-R models, adaptive D-R studies, and the original MCP-Mod method, respectively. In Section 2.4, a conditional error function, which is used to combine two-stage information, is reviewed. In Sections 3 and 4, our extended MCP-Mod method and simulation studies are given. Finally, in Section 5, we discuss our method and results.

2 Background

2.1 Examples of Dose-Response Models

We now review important dose-response models. Let $f(d, \theta)$ denote a general steady-state D-R model with a parameter vector θ , which is used to describe a mean response at a certain

dose level d . In the MCP-Mod method, a standardized version of the D-R models is used. In the corresponding *standardized* models $f^0(\cdot)$, the parameter vector is denoted by θ^* .

The $E_{\max}$ model is a simple empirical D-R model (Dorendorf and Meibohm, 1999), and its general form is

$$f(d, \theta) = E_0 + \frac{E_{\max} d}{EC_{50} + d} \quad \text{with } \theta = (E_0, E_{\max}, EC_{50}). \quad (1)$$

In (1), E_0 is a baseline effect, $E_{\max}$ is the maximum change associated with d , and EC_{50} is the dose that produces half of the maximum change (Branson, Pinheiro, and Bretz, 2003).

Logistic models and log-linear models are also commonly used to study a relationship between a drug concentration and a drug effect (Meibohm and Derendorf, 1997). Branson, Pinheiro, and Bretz (2003) used the following logistic model in their development of the MCP-Mod approach:

$$f(d, \theta) = E_0 + \frac{E_{\max}}{1 + \exp\left(\frac{EC_{50} - d}{\delta}\right)} \quad \text{with } \theta = (E_0, E_{\max}, EC_{50}, \delta), \quad (2)$$

where E_0 is the baseline effect, $E_{\max}$ is the maximum effect size from the baseline effect, EC_{50} is the dose producing half of $E_{\max}$, and δ is the parameter related to the rate of response change with dose. In (2), E_0 and $E_{\max}$ are location and scale parameters, respectively. Another possible model choice is the log-linear model (Branson et al., 2003),

$$f(d, \theta) = E_0 + \delta \log(d + c), \quad (3)$$

where $\theta = (E_0, \delta)$, and c is a constant which satisfies the inequality, $c > -d$ for all d . We summarize specifications of the original and standardized $E_{\max}$, quadratic, and logistic models and their parameters in Table 3. Figure 3 gives their general shapes.

2.2 Adaptive Dose-Response Studies

Adaptive D-R studies allow one to study evidence of treatment effect and characterize the relationship between dose and response, while using interim data to decide how to modify design features for refining the studies. According to PISC of PhRMA (Bornkamp et al., 2007), the general objectives of adaptive D-R studies are: (1) to show proof of concept; (2) to establish the D-R curve; (3) to select the optimal dose; and (4) to estimate the therapeutic window. For these objectives, various approaches for a D-R study have been developed. Typical examples include 1) a Bayesian adaptive dose allocation approach (Berry et al., 2002), 2) a D-optimal response-adaptive approach (Dragalin and Fedorov 2006), 3) a multiple comparison procedures-modeling approach (MCP-Mod) (Branson et al., 2003 and Bretz et al., 2005), 4) a Bayesian model-averaging approach (Bornkamp et al., 2007), 5) a multiple trend test approach (Bornkamp et al., 2007), and 6) a non-parametric D-R modeling approach (Bornkamp et al., 2007). In this paper, we chose the MCP-Mod method—a non-adaptive design with frequentist adaptive analysis—as basis for proposing a two-stage adaptive *design* method.

2.3 Multiple Comparison Procedures-Modeling (MCP-Mod) Method

The MCP-Mod method (Branson et al., 2003 and Bretz et al., 2005) addresses three kinds of decision-making within a dose-finding study: (i) show proof of concept (PoC) that the treatment is effective, (ii) obtain the best D-R relationship, and (iii) choose dose(s) to use in further development. In the establishment of PoC, a set of candidate models which cover a suitable range of D-R shapes is used instead of a single pre-specified D-R model. The MCP-Mod method has two steps: the MCP step (multiple comparison procedures) and the Mod step (modeling). If PoC has been established, the best model is selected (Mod) from the statistically significant models in the candidate set. Furthermore, the selected D-R model is used to estimate a target dose for clinical use. We now give a more detailed description of the MCP step.

The MCP step is necessary for the establishment of PoC. It is assumed that responses are observed for a given set of $k + 1$ parallel groups of patients randomized to placebo denoted as d_0 , and k treatment doses, $d_1, d_2, \dots, d_k$. Consider the D-R model $y_{ij} = f(d_i, \theta) + \epsilon_{ij}$ for the observed data where y_{ij} is the response of the subject j in the i^{th} group of size n_i , d_i is the dose the subject receives, $f(d_i, \theta)$ is the mean response of the group given a certain D-R model, $f(\cdot)$, with a parameter vector, θ , and ϵ_{ij} is the independent homoscedastic normal error. Using *optimal contrast test statistics*, one can decide which D-R models give us statistically significant fits. The test statistics are defined by *optimal contrasts*. To understand such statistics, suppose that one pre-specifies a set $\mathcal{M}$ of M parameterized candidate models, $f_m(d_i, \theta)$, $m = 1, \dots, M$, and that each model is standardized. Each D-R shape can be tested by a single contrast test associated with optimized coefficients obtained from each standardized model. These optimized coefficients (optimal contrasts) are chosen in advance to maximize the power of the test when the true underlying mean response at each dose is the modeled value. Specifically, the single contrast tests are given by

$$T_m = \frac{\sum_{i=0}^k c_{mi} \bar{y}_i}{S \sqrt{\sum_{i=0}^k \frac{c_{mi}^2}{n_i}}}, \quad m=1, \dots, M, \quad (4)$$

where $S^2 = \frac{\sum_{i=0}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{N - k - 1}$, $N = \sum_{i=0}^k n_i$, and $c_{m0}, \dots, c_{mk}$ are the optimal contrasts

for model m . The contrasts are chosen to satisfy $\sum_{i=0}^k c_{mi} = 0$ and $\sum_{i=0}^k \frac{c_{mi}^2}{n_i} = 1$ and maximize the centrality parameter of the contrast test for the standardized model m (see Branson et al. and Bretz et al.). The MCP to test whether or not there is a dose-response is based on the maximum of $T_1, \dots, T_M$ via these standardized models. To control the FWER, the multiplicity adjusted p-value and critical values can be numerically computed based on the assumption that the multiple test statistics, $(T_1, \dots, T_M)'$, are jointly multivariate t -distributed under the null hypothesis of no dose-response with correlation matrix determined by the model contrasts. One accepts or rejects the null hypothesis based on the adjusted p-value at a pre-chosen significant level α . Rejecting the null hypothesis establishes the PoC. The MCP is useful because it accounts for the uncertainty of model specifications at the beginning of the trial.

Once the PoC is established, the model and dose selection is performed in the Mod step. After the controlled PoC tests, one takes statistically significant models from the candidate model set, $\mathcal{M}$. The best model is selected from these significant models based on its having the minimum p-value of the contrast test statistics. Other model selection criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) can be also used.

2.4 Testing Global Null Hypothesis in Two-Stage Designs

Employing the adaptive test procedure by Bauer and Köhne (1994), Bretz et al. (2006) conceptualized the use of combination tests in the ‘seamless’ Phase II/III adaptive trials where a conditional error function (CEF) is applied. ‘Seamless’ is usually used to describe the adaptive designs that combine two trials of two phases into one protocol. A CEF, $A(p_1)$, is defined as the probability of rejecting a global null hypothesis, H_g , given the p-value of Stage 1, p_1 , in a two-stage adaptive design.

A decision rule for rejecting the global null hypothesis can be formulated using the CEF of the combination test as follows:

$$\begin{aligned} &\text{reject } H_g \text{ if } p_2 < A(p_1); \\ &A(p_1) = P_H(\text{reject } H | p_1) = \begin{cases} 1 & \text{if } p_1 \leq \alpha_1 \\ 0 & \text{if } p_1 \geq \alpha_0 \\ \max\{p_2 | C(p_1, p_2) \leq \mathcal{C}_\alpha\} & \text{if } p_1 \in (\alpha_1, \alpha_0) \end{cases} \quad (5) \end{aligned}$$

where p_1 and p_2 are the p-values from testing the null hypotheses in Stage 1 and Stage 2, respectively. Also, α_0 and α_1 are the stopping boundaries at the interim analysis for futility and efficacy, respectively, C is a function which combines test statistics across the two stages, and $\mathcal{C}_\alpha$ is a pre-chosen critical value at some α level.

The above CEF, $A(p_1)$, is useful for combining any kind of statistics when one specifies the function form of its combination test, $C(p_1, p_2)$. In our combining of the two stages, we employ Fisher’s combination test as discussed in Bauer and Kieser (1999). Because we test the same set of pre-specified candidate models in both stages, their assumptions to control Type I error are met. Fisher’s combination test is given by $C(p_1, p_2) = p_1 p_2$ and

$\mathcal{C}_\alpha = \exp(-\chi_{4,1-\alpha}^2)/2$. This test assumes that both p_1 and p_2 follow the uniform distribution (0, 1) under the null hypotheses in Stage 1 and Stage 2, respectively. The general concept of Bretz et al. can be applied for a global establishment of PoC in a two-stage dose adaptive design with an adding and/or dropping treatment adaptation rule as well.

3 Proposed Extended MCP-Mod Method (xMCP-Mod)

The goal of our approach is to establish global PoC in the adaptive two-stage design. With the exception of adding and/or dropping doses, all potential doses are pre-specified. The result of the first stage allows one to select the doses for the second stage via a pre-specified adaptation rule (which we call ADTAR). The global result is obtained by combining the results of the PoC test in Stage 1 and the adaptive PoC test in Stage 2 using Fisher’s

combination test. Possible sample size imbalances in the second stage due to the adaptation rule are addressed through a weighting scheme which will be described in Section 3.5.

3.1 Overview of xMCP-Mod

In our ADTAR study design, candidate dose combinations for Stage 2 are pre-specified and one combination is selected after Stage 1. The set of possible rules is formally characterized. This pre-specification follows a recommendation of the FDA's draft guidance (FDA 2010) and considers regional non-U.S. regulatory views. From a statistical point of view, ADTAR allows us to prospectively formulate all candidate null hypotheses in the protocol for the PoC test of Stage 2, and we adaptively select one null hypothesis for Stage 2 after Stage 1 is performed. Note that every adaptation choice is associated with a separate null hypothesis since multiple contrasts, distributions, and critical values depend on dose adaptation choices. With the adding and/or dropping dose adaptation, we aim to improve design power in the global PoC test over conventional (one-stage) and fixed two-stage designs. In contrast, fixed two-stage designs are similar to designing two independent studies. The data collected for establishing PoC can be generally used for estimating the best dose-response relationship and finding an optimal dose, using one of the existing methods once the PoC is established. For instance, the original (one-stage) MCP-Mod method can be applied to the full data set.

The framework of the adaptive two-stage design is described as follows. In each stage, we respectively apply the MCP-Mod method of Bretz et al. The two stages share the same candidate D-R models and total sample size. In each stage, the total sample size is equally distributed across dose groups. The only difference of Stage 2 from Stage 1 is the choice of doses, which is determined through a yes/no decision for the individual pre-specified potential doses in the adaptation process. In each stage, a PoC test is performed using contrast test statistics as done in the MCP-Mod method. We additionally utilize each model-based contrast statistic in Stage 1 for constructing an indicator vector. This indicator vector shows whether or not the candidate D-R models are preliminarily significant. The indicator vector will be used in the dose adaptation for Stage 2.

Using the ADTAR rule, the result of the Stage 1 PoC test will determine what doses and tests will be used in Stage 2. This is done through ADTAR, a function that maps Stage 1 data to the doses to use in Stage 2 deterministically and is pre-specified before the trial is initiated. This rule maps any model-significance indicator vector from Stage 1 into a dose specification in Stage 2. Depending on the pre-specified doses in Stage 1 and in the selected adaptation choice, the new yes/no vector of doses for Stage 2 effectively adds and/or drops doses with respect to Stage 1. It should be essential for trial designers to account for existing pre-clinical and clinical data in optimizing an adaptation rule for a particular investigational drug.

When obtaining model-based contrast test statistics for Stage 2, we construct a separate null hypothesis for each adaptation choice of the rule, and the p-values derived at the null hypothesis tests are computed with respect to the multivariate t-distribution and critical value of the adaptation choice. In this way, two sources of dependence of stages (multiple contrast coefficients and critical value) are resolved. Furthermore, we make adjustments that account for the following factors. First, not all the doses appear in every adaptation choice.

Second, a fixed total number of samples have to be distributed across the doses in a particular Stage 2. Consequently, the number of samples per dose is data dependent and cannot be known at design time. We model this data dependency as a missing data mechanism and compensate for the imbalance using a weighting method from missing data analysis (Little and Rubin 2002). This approach models the distribution of missing data and derives weights that adjust for biases of estimates. We are using these weights to adjust the contrast test statistics of Stage 2.

Following the MCP-Mod method, p-values are obtained for Stage 1 and Stage 2 based on a multivariate t -distribution. The overall or global null hypothesis is that there is no dose-response relationship in the adaptive two-stage trial. The global PoC test is performed by combining the two p-values using Fisher's combination test.

We formulate our xMCP-Mod method to maximize the power for detecting PoC assuming that a separate PoC establishment is an a priori requirement due to a corporate management decision or a regional regulatory view before expending resources on estimating the optimal dose. The power for establishing PoC of such a design method is not directly comparable with the power of other adaptive designs with different objectives. After establishing PoC, an optimal dose still needs to be estimated. This could be done, for instance, by applying the MCP-Mod method to the full data set. In such a case, the result may suffer from the dose allocation designed for establishing PoC. The ideal solution would be to extend our method to handle the remaining steps after establishing PoC, which is a topic of our future research.

3.2 Example trial

To better illustrate our approach we consider a hypothetical example that we use throughout the paper including the simulation study.

Procedure

1. Pre-specification

Prior to initiating the trial, we pre-specify the following items. Our hypothetical study design is described below.

- The equal total sample size in both stages (40, 100, 200, or 300 subjects in each stage are employed in simulations).
- Significance level: throughout the example we use 0.05 as threshold for statistical tests.
- Three standardized candidate D-R models (see Section 2.1): an $E_{\max}$ model denoted as f_1 with its parameter $ED_{50} = 0.2$, a quadratic model denoted as f_2 with $\delta_{\max} = 0.6$, and a logistic model denoted as f_3 with $ED_{50} = 0.2$ and $\delta = 0.6$. In a real trial, these parameters would be derived from prior pre-clinical and clinical knowledge.
- Placebo $d_0 (= 0)$ and four potential doses $(d_1, \dots, d_4) = (0.05, 0.2, 0.6, 1.0)$. The four doses could potentially be used in Stage 1 and any Stage 2 adaptation choices.

- A subset of the potential doses that is used in Stage 1: (d_0, d_1, d_2, d_3) .
- An adaptation rule: As a result of Stage 1 that drives the adaptation for Stage 2, we use an indicator vector $\mathbf{b} = (f_1 \text{ is significant}, f_2 \text{ is significant}, f_3 \text{ is significant})$ to indicate Stage 1 result. This vector is in $\{0, 1\}^3$; there are eight possible vectors. To select a set of doses from the pre-specified adaptation choices for Stage 2, the rule takes the indicator vector $\mathbf{b}$ and returns a yes/no decision for each dose $d_0, \dots, d_4$, whether it will be included. We call this particular example of an adaptation rule ADTAR-rule, shown in Table 1.

In Table 1, the number of doses considered in Stage 2 can be different among the adaptation choices, with the number of subjects equal across doses. Because the equal total sample size is the same for both stages, sample sizes per dose group is not known in advance. Therefore, sample sizes per dose can be unbalanced across multiple trials of the same design if the ADTAR-rule does not select doses uniformly across adaptation choices. (Suppose total stage-wise sample size is 200. If an adaptation choice specifies four doses, the sample size per dose group in Stage 2 is 50. If five doses are specified in another adaptation choice, the sample size per dose group in Stage 2 is 40.) This imbalance adds complication for test statistics in Stage 2 as the sample size per dose is not known at design time.

2. Stage 1

Stage 1 is performed using the design specification as described above. In the interim statistical analysis following Stage 1, the PoC test of the MCP-Mod method is used to obtain the p-value based on the multivariate the t -distribution. This p-value will be utilized to compute the global p-value after Stage 2 is performed. Furthermore, p-values based on univariate t -distributions are obtained and thresholded with the pre-specified confidence level to produce the indicator vector $\mathbf{b}$. This indicator vector is used in the adaptation as described next.

3. Adaptation

- We use the indicator vector to find the set of doses for Stage 2 given the preliminary result of Stage 1. For instance, if the three model-associated p-values based on the univariate t -distributions, p_1 , p_2 , and p_3 , had values of 0.10, 0.01, and 0.27, respectively, we would conclude that the quadratic model is a significant model while the other Emax and logistic models are not significant using the 0.05 significance level. Based on this interim result, $\mathbf{b} = (0, 1, 0)$. The pre-specified rule ADTAR-rule (Table 1, the fourth row in the column *Model Significance*) yields corresponding adaptation choice 2 (the second row in the right column) where we drop dose d_3 and add dose d_4 , thus, selecting d_0 , d_1 , d_2 , and d_4 as the dose adaptation choice to use in Stage 2.
- If stage-wise sample size=200, based on the noted Stage 1 outcome, there would be 50/dose group in Stage 2.

- Given the doses and group size for Stage 2, we compute weights to use for the contrast statistics in Stage 2. The rationale and method to compute the weights will be described in Section 3.5. All possible weights for doses corresponding to the adaptation choices 1–4 in Table 1 are shown in Table 2. Each column contains the weights for a particular adaptation choice. In our example, we selected the second adaptation choice which yields the weights $(1/4, 1/3, 1/3, 0, 2/7)$ as shown in the second column in Table 2.

4. Stage 2

- Stage 2 is performed using the adaptively selected doses and the corresponding group sizes. In our case, we use the selected doses d_0, d_1, d_2 , and d_4 .
- The statistical test for Stage 2 PoC follows the MCP-Mod idea but applies our weighting procedure to compensate for the sample size imbalance. The optimal contrast vector according to MCP-Mod is weighted by the weights from Table 2, and the p-value for PoC is based on the weighted multivariate t -distribution.

5. Global PoC

- The statistical test for the overall or global PoC test is obtained by combining the previously obtained Stage 1 and Stage 2 multivariate t p-values using Fisher's combination test.

3.3 Adding and/or Dropping Treatment Adaptation Rule (ADTAR)

In this section, we formally define the adding and/or dropping treatment adaptation rule (ADTAR). This is a core contribution of the paper. In our trial design, there are M candidate D-R models, $f_1, \dots, f_M$. As discussed in Section 2.1 and used in our trial example (Section 3.2), some possible candidate models are: Emax, quadratic, and logistic models. The M candidate D-R models are standardized into $f_1^0, \dots, f_M^0$ following the standardization technique described in Section 2.3. Prior to the trial, we also pre-specify placebo d_0 and k potential doses, $d_1, \dots, d_k$. In our example, we used placebo and four potential doses. Throughout Stages 1 and 2, no other dose will be used but not all k doses *have* to be used. In our example and simulation, we chose normalized doses between 0 and 1 to make our xMCP-Mod method comparable to the MCP-Mod method of Bretz et al. For Stage 1, we pre-specify a subset of the k doses as treatment doses to be used. In our example, we used the first three treatment doses as the subset. In Stage 2, treatment doses will be adaptively selected from the k doses based on the ADTAR rule. A placebo group, d_0 , is always used throughout the trial.

Formally, an ADTAR rule is a function that maps the interim or preliminary result of Stage 1 into a dose choice of Stage 2. From the interim data of Stage 1, we obtain M model-associated p-values based on M univariate t -distributions. In our trial example, we had p_1, p_2 , and p_3 . We thresholded these p-values using a pre-specified significance level α (in our case, 0.05) to determine whether or not each model is significant at the interim analysis. This leads to a M -dimensional indicator vector

$$\mathbf{b} = (I(\text{p-value for } f_1 < 0.05), \dots, I(\text{p-value for } f_M < 0.5))$$

that encodes the $\mathcal{K}_1 = 2^M$ possibilities of combinations for significance of the M D-R models in Stage 1. $I(\cdot)$ denotes the indicator. Since there were three D-R models in the trial example, we listed $2^3 = 8$ possible combinations in the left column of Table 1. Note that this particular example of an adaptation rule in Table 1 is not optimized. Doses tend to be added when they are likely to differentiate two or three D-R models. However, this is not always the case, and we also included an adaptation choice where there is no dose to add or drop to show more general possibilities enabled by our approach. We artificially crafted the rule to illustrate the ADTAR approach.

The vector $\mathbf{b}$ is the input to the adaptation function g ; the set $B = \{0, 1\}^M$ is the domain of the ADTAR function g . We denote the elements of B as $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{\mathcal{K}_1}$, using their canonical lexicographic order¹ to number them. For our p-values, $p_1 = 0.10$, $p_2 = 0.01$, and $p_3 = 0.27$, the indicator vector was $\mathbf{b} = (0, 1, 0)$.

The output of the adaptation rule is the set of doses to be used in Stage 2. This is given as an indicator vector,

$$\mathbf{a} = (I(d_0 \text{ is used}), I(d_1 \text{ is used}), \dots, I(d_k \text{ is used})).$$

In Stage 2, because the possible adaptation with no non-placebo treatment doses is not allowed, there are $\mathcal{K}_2 = 2^k - 1$ possible sets of doses that could be examined. In our trial example, since we had $k = 4$ treatment doses, $\mathcal{K}_2 = 15$ sets of doses could have been examined in Stage 2.

We do not necessarily allow all $\mathcal{K}_2$ dose combinations as adaptation choices. Accordingly, we may specify only L adaptation choices for Stage 2 where $L \leq \mathcal{K}_2$. Thus, the rule implicitly pre-specifies L binary vectors, $\mathbf{a}_j$, each of dimension $k + 1$, defining whether or not each dose d_i , $i = 0, \dots, k$ is included in the j^{th} adaptation choice, $j = 1, \dots, L$. The set of L adaptation choices is denoted by $A = \{\mathbf{a}_1, \dots, \mathbf{a}_L\}$ and this set is the range of the adaptation function g . In the trial example, the number of rows in the right column of Table 1 was $L = 4$. Each row represented the binary vector $\mathbf{a}_j$, $j = 1, \dots, 4$. Therefore, we had the set $A = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ with $\mathbf{a}_1 = (1, 0, 1, 1, 1)$, $\mathbf{a}_2 = (1, 1, 1, 0, 1)$, $\mathbf{a}_3 = (1, 1, 1, 1, 0)$, and $\mathbf{a}_4 = (1, 1, 1, 1, 1)$.

We now formally define the adaptation function g , called *ADTAR rules*, $g : B \rightarrow A$ and $\mathbf{a} = g(\mathbf{b})$, as a mapping of the M -dimensional indicator vector $\mathbf{b}$ into a $k + 1$ -dimensional indicator vector $\mathbf{a}$. Define B_j as the subset of B that maps to the adaptation choice $\mathbf{a}_j$, $B_j = g^{-1}(\mathbf{a}_j)$,² i.e., B_j , $j = 1, \dots, L$, partitions the set B into L elements according to the L adaptation

¹The lexicographic order of vectors is defined as ordering them dimension by dimension. Vectors are ordered according to their first dimension; vectors with the same value in the first dimension are ordered according to the second dimension, and so forth. For example, $(0, 5, 1) < (1, 1, 4) < (1, 2, 3)$.

²We use $g^{-1}(\mathbf{a})$ as shorthand for the set $\{\mathbf{b} \in B : g(\mathbf{b}) = \mathbf{a}\} \subseteq B$, i.e., the set of all elements $\mathbf{b}$ of B that the function $g(\cdot)$ maps to $\mathbf{a}$.

choices. In Table 1 of the trial example, the rows in the left column were grouped into one of subsets B_1 , B_2 , B_3 , and B_4 given $L = 4$ adaptation choices. The first and second rows on model significance, $(1, 1, 1)$ and $(1, 1, 0)$, were grouped into subset B_1 . Similarly, the third and fourth rows, $(0, 1, 1)$ and $(0, 1, 0)$, were grouped into B_2 , the fifth row itself, $(0, 0, 1)$, made B_3 , and the rest of the rows, $(1, 0, 1)$, $(1, 0, 0)$, and $(0, 0, 0)$, were grouped into B_4 .

To fully define g , we need to specify the sets $B_1, \dots, B_L$ and $\mathbf{a}_j$ corresponding to each B_j . This defines which Stage 1's outcome maps to which adaptation choice for Stage 2. We give a closed form expression for g . To explicitly obtain a particular $\mathbf{a}$ by computing $g(\mathbf{b})$, we define the ADTAR design matrix,

$$\mathbf{Z}_{L,k+1} = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_L \end{pmatrix}.$$

We also define an indicator vector

$$\mathbf{v}(\mathbf{b}) = (I(\mathbf{b} \in B_1), I(\mathbf{b} \in B_2), \dots, I(\mathbf{b} \in B_L)), \quad (6)$$

which is a L -dimensional basis vector. Then, we can compute the $k + 1$ -dimensional adaptation outcome vector $\mathbf{a}$ by

$$\mathbf{a} = \mathbf{v}(\mathbf{b}) \cdot \mathbf{Z}_{L,k+1}. \quad (7)$$

As in the previous example, we used $\mathbf{b} = (0, 1, 0)$ for the Emax, quadratic, and logistic models indicating that only the quadratic model was significant in the interim analysis of Stage 1. Looking at Table 1, the fourth row in the left column, $(0, 1, 0)$, was the input vector $\mathbf{b}$. Table 1 mapped the fourth row in the left column to the second row in the right column among four adaptation choices, since $\mathbf{b} = (0, 1, 0) \in B_2 = \{(0, 1, 1), (0, 1, 0)\}$. Using (6), our 4-dimensional basis vector was $\mathbf{v}(0, 1, 0) = (0, 1, 0, 0)$. Since the right column of Table 1 represented the design matrix of ADTAR-rule with $k = 4$ potential treatment doses,

$$\mathbf{Z}_{4,5} = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

we computed the adaptation choice associated with $\mathbf{b} = (0, 1, 0)$ as

$$\mathbf{a} = \mathbf{v}(0, 1, 0) \cdot \mathbf{Z}_{4,5} = (0, 1, 0, 0) \cdot \mathbf{Z}_{4,5} = (1, 1, 1, 0, 1).$$

Thus, the selected dose combination for Stage 2 based on the interim outcome, $\mathbf{b} = (0, 1, 0)$, was the second adaptation choice $\mathbf{a}_2 = (1, 1, 1, 0, 1)$ where we use d_0 , d_1 , d_2 , and d_4 but not d_3 in Stage 2.

Equation (7) provides the formal definition of ADTAR rules. An ADTAR rule is fully characterized by a binary $L \times k + 1$ matrix where the first column is constant 1 and all rows are different, and a partition of B into $\cup_{j=1}^L B_j$. This characterization allows enumeration of all possible ADTAR rules and further calculation based on them as in Section 3.5.

3.4 Test Statistics of Stage 1

Test statistics for Stage 1 can be obtained using the same procedure as in the MCP-Mod method. Recall from Section 2.3 that the original MCP-Mod method assumed that the multiple test statistics, $(T_1, \dots, T_M)'$ as defined in 2.3, are jointly multivariate t -distributed under the null hypothesis of no dose-response. Let $(T_{1,1}, \dots, T_{1,M})'$ denote a vector of the contrast test statistics in Stage 1. Further, let MVT_{M, ν_1} denote a M -dimensional multivariate t -distribution with ν_1 degrees of freedom. Recall that $\nu_1 = N_1 - k_1 - 1$ where N_1 is the total sample size in Stage 1 and k_1 is the number of treatment doses in Stage 1. Then, under the null hypothesis of Stage 1, $(T_{1,1}, \dots, T_{1,M})' \sim MVT_{M, \nu_1}(\mathbf{0}, \mathbf{R}_1)$ with a correlation matrix, $\mathbf{R}_1 = (\rho_{uv})_1$, where the matrix elements are given by

$$\rho_{uv,1} = \frac{\sum_{i=0}^{k_1} \frac{c_{ui}c_{vi}}{n_{1i}}}{\sqrt{\sum_{i=0}^{k_1} \frac{c_{ui}^2}{n_{1i}}} \sqrt{\sum_{i=0}^{k_1} \frac{c_{vi}^2}{n_{1i}}}}, \quad 1 \leq u, v \leq M,$$

and the optimal contrast vector, $\mathbf{c}_m = (c_{m0}, \dots, c_{mk_1})'$.

3.5 Test Statistics of Stage 2: Missing Data Analysis

The derivation of the test statistics of Stage 2 is based on the MCP-Mod method, but we need to compensate for the impact of adaptation on sample sizes. Stage 2, seen in isolation, is a study using the MCP-Mod method, and viewed as such, the test statistics and optimal contrast vectors are derived using the same procedure as in Stage 1. All potential Stage 2 designs are defined through a pre-selected dose combination according to an adaptation choice $\mathbf{a}_j \in A$. For a particular $\mathbf{a}_j$, the test vector of an isolated Stage 2 under the null hypothesis of no dose-response, $(T_{2,1}, \dots, T_{2,M} | \mathbf{a}_j)' \sim MVT_{M, \nu_2}(\mathbf{0}, \mathbf{R}_2)$ with a correlation matrix, $\mathbf{R}_2 = (\rho_{uv})_2$, where the matrix elements are given by

$$\rho_{uv,2} = \frac{\sum_{i=0}^{k_2} \frac{c_{ui}^* c_{vi}^*}{n_{1i}}}{\sqrt{\sum_{i=0}^{k_2} \frac{c_{ui}^{*2}}{n_{2i}}} \sqrt{\sum_{i=0}^{k_2} \frac{c_{vi}^{*2}}{n_{2i}}}}, \quad 1 \leq u, v \leq M,$$

where $\mathbf{c}_m^* = (c_{m0}^*, \dots, c_{mk_2}^*)'$ is the optimal contrast vector which depends on the pre-selected doses of the particular Stage 2 design, as mandated by $\mathbf{a}_j$.

Our ultimate goal is to combine the results of the null hypothesis tests in Stage 1 and Stage 2 into a global null-hypothesis test, using Fisher's combination test. Fisher's combination test requires Stage 1 and Stage 2 to be statistically independent. However, Stage 2 is dependent on Stage 1 because Stage 2 is parameterized by $\mathbf{a}_j$ that was selected based on the Stage 1

outcome. Since we construct a separate null hypothesis for each adaptation choice of the rule, and the p-values derived from these null hypothesis tests are computed with respect to the multivariate t -distribution and critical value of the adaptation choice. In this sense the p-value in Stage 2 is independent of the adaptation choice. There is one source of dependency from data of Stage 1 that cannot be treated this way. Since the total sample size of Stage 2 is constant, the number of samples per dose depends on the chosen adaptation choice. This results in the means per dose in different adaptation choices being computed from different sample sizes, depending on the adaptation choice. Consequently, the PoC contrast test statistic needs to be adjusted across adaptation choices. As we discuss next, we use a missing data analysis approach to compensate for this dependency and derive a new test statistic for Stage 2 that makes Stage 2 statistically independent from Stage 1 for testing the null hypothesis.

In summary, Stage 2 depends on the ADTAR function g which uses the result from Stage 1 to construct the adaptation choice. Across multiple trials, and depending on g , some doses may not be sampled in each trial's Stage 2. In our setting, we can view the adaptation for Stage 2 as creating a missing data mechanism (Little and Rubin 2002). Accordingly, the response estimate for dose i can be adjusted by applying a missing data analysis method. The missing data mechanism within Stage 2 is Missing At Random (MAR) because the missingness with respect to the doses depends only on the observed part of the $\mathcal{X}_1$ potential PoC results in Stage 1. An example illustrating the missing data mechanism is shown in Section 4.3. Since the adaptation choice determines which doses are 'missing', we can then condition on that choice, rendering the missing data mechanism as Missing Completely At Random (MCAR), i.e., quasi-randomized (Little and Rubin 2002). Under the assumption of quasi-randomization, one can use the weighting class estimator of missing data analysis. This unbiased estimator uses the idea that the observations associated with adaptation choice, $\mathbf{a}_j$ (called 'stratum j ' by Little and Rubin), receive a weight for each dose group, i . We let w_{ij} denote the weight which corresponds to the mean estimate of dose i for adaptation $\mathbf{a}_j$. In Appendix A, we derive the weights, w_{ij} .

Using our derived weights w_{ij} , we define the distribution of the test statistics for a given adaptation choice $\mathbf{a}_j$ under the null hypothesis of Stage 2. The derived weights adjust the test statistic of each dose group in adaptation choice $\mathbf{a}_j$ and thus resolve the dependency of Stage 2, as detailed in Appendix A. The adjusted statistic on model m for adaptation $\mathbf{a}_j$ in Stage 2 is

$$T_{2,m}^{\mathbf{a}_j} = \frac{\sum_{i=0}^{k_2} w_{ij} c_{mi}^* \bar{y}_{ij}}{S_2 \sqrt{\sum_{i=0}^{k_2} \frac{(w_{ij} c_{mi}^*)^2}{n_{2i}}}}, \quad m=1, \dots, M,$$

where $S_2^2 = \frac{\sum_{i=0}^{k_2} \sum_{h=1}^{n_{2i}} (y_{ij} - \bar{y}_{ij})^2}{N_2 - k_2 - 1}$ and $\nu_2 = N_2 - k_2 - 1$. In summary, the independent distribution of Stage 2's test statistics for choice $\mathbf{a}_j$ is given by

$$(T_{2,1}^{\mathbf{a}_j}, \dots, T_{2,M}^{\mathbf{a}_j})' \sim \text{MVT}_{M, \nu_2}(\mathbf{0}, \mathbf{R}_2^*) \text{ where } \mathbf{R}_2^* = (\rho_{uv}^*)_{2^2}$$

$$\rho_{uv,2}^* = \frac{\sum_{i=0}^{k_2} \frac{w_{ij}^2 c_{ui}^* c_{vi}^*}{n_{2i}}}{\sqrt{\sum_{i=0}^{k_2} \frac{w_{ij}^2 c_{ui}^{*2}}{n_{2i}}} \sqrt{\sum_{i=0}^{k_2} \frac{w_{ij}^2 c_{vi}^{*2}}{n_{2i}}}},$$

and $1 \leq u, v \leq M$. By using the above weighted statistics and correlations, we compensate the effect of unbalanced sample allocation due to ADTAR.

3.6 Combining Stage 1 and Stage 2

In the last step, we form the multiple contrast test statistics and obtain a p-value in each stage by use of multivariate t -distributions, i.e., $(T_{1,1}, \dots, T_{1,M})' \sim \text{MVT}_{M, \nu_1}(\mathbf{0}, \mathbf{R}_1)$ in Stage 1 and $(T_{2,1}^{a_j}, \dots, T_{2,M}^{a_j})' \sim \text{MVT}_{M, \nu_2}(\mathbf{0}, \mathbf{R}_2^*)$ in Stage 2 using the original MCP-Mod method (Branson et al., 2003 and Bretz et al., 2005). Let p_1 and p_2 denote the obtained p-values in Stage 1 and Stage 2, respectively. In the global PoC test, we plug these p-values into the CEF formula from the combination technique (Section 2.4) to test the global null hypothesis

of no dose-response relationship. The global PoC is established if $p_2 < A(p_1) = \frac{\mathcal{C}_\alpha}{p_1}$ where $\mathcal{C}_\alpha$ is a pre-chosen critical value.

4 Simulation Study: Evaluating xMCP-Mod

In this section, we evaluate the performance of our adaptive design method, xMCP-Mod. Since we apply our method for detecting PoC, we assess its performance by measuring probability of detecting a true D-R relationship. To assess whether an ADTAR rule has an impact when combining two stages, we also measure the FWER using an non D-R relationship (a constant model).

First, we compare the xMCP-Mod method with the original MCP-Mod design and aim to show the improvement achieved through adding and/or dropping treatment arms in the second stage based on the preliminary result of the first stage. We also compare the xMCP-Mod method with four different fixed two-stage designs to investigate the impact of adaptivity versus adding and/or dropping doses in a fixed way. Furthermore, we investigate the impact of the missing data weighting procedure of ADTAR for compensating the sample size imbalance of the second stage against the same ADTAR design without weighting. Lastly, to further investigate the impact of ADTAR on controlling the FWER we compare the xMCP-Mod method with a random two-stage design which has 15 equally possible treatment arms in the second stage.

4.1 Simulation Setup

Below we use the notation described in Section 3 when discussing parameters of the xMCP-Mod method.

- **Study designs:** Figure 1 illustrates the three designs we evaluate for probability of detecting PoC (PoC probability): (a) four fixed two-stage designs, (b) our proposed adaptive two-stage xMCP-Mod design, and (c) a one-stage design according to the

original MCP-Mod method. For the FWER assessment we use a random two-stage design that uniformly selects the second stage out of all the 15 possible dose combinations (see Figure 2).

- **Performance measurement:** All data is based on simulating 10,000 trials. Overall performance was measured by quantifying the probability of detecting a D-R curve in 16 different scenarios, which were the combinations of four different models (Emax, quadratic, logistic, and constant models) and four different total sample sizes ($N = 80, 200, 400$, and 600) for each study design. When evaluating the xMCP-Mod, the equal sample sizes for group i in Stage 1 were $n_{1i} = 10, 25, 50$, and 75 . In all the adaptations, balanced sample sizes for each group in Stage 2 (n_{2i}) were chosen so that the total number of observations between Stage 1 and Stage 2 would be the same. To assess the FWER we calculated probabilities of detecting a significant D-R curve under the null hypothesis of a constant model.
- **Candidate D-R models:** Three standardized candidate D-R models were used ($M = 3$) for both the original MCP-Mod method and our xMCP-Mod method: Emax, quadratic, and logistic models. The model specifications are summarized in Table 3 and Figure 3 where $ED_{50} = 0.2$ for f_1^0 , $\delta_{max} = 0.6$ for f_2^0 , and $ED_{50} = 0.2$ and $\delta = 0.6$ for f_3^0 . These models were used to compute the optimal contrast vectors.
- **Data generation:** We generated data using the same models that were used as candidate D-R models in both MCP-Mod and xMCP-Mod. In particular, four datasets were generated based on a population mean vector of constant, Emax, quadratic, and logistic models, respectively. All the models had the property that a maximum effect size within the interval $[0, 1]$ was 0.8 and that the response at the placebo was 0.2 . The constant data-generating model was $f(d_i) = 0.2$. We assumed independent outcomes from $N(f(d_i), \sigma^2)$ at d_i where $\sigma = 1.478$.
- **Doses:** Throughout all experiments we used the doses $(d_0, d_1, d_2, d_3, d_4) = (0, 0.05, 0.2, 0.6, 1.0)$. Thus, the number of potential doses for the xMCP-Mod method was four ($k = 4$).
- **Significance level:** We employed a 2.5% one-sided significance level to achieve 95% confidence intervals in all PoC tests. This is the same confidence interval used in Bretz et al. (2005). The mismatch between our significance level and confidence interval is due to the fact that we used a one-sided test in order to perform the necessary symmetric two-sided test (Branson et al., 2003).
- **Adaptation in xMCP-Mod (ADTAR function):** Throughout all experiments we used one particular ADTAR function as the illustrative example. We called this adaptation function ‘ADTAR-rule’. The function was informally used in Section 3.2; we here fully define all parameters. Table 4 presents all the parameters. In xMCP-Mod’s Stage 1, the dose vector $(d_0, d_1, d_2, d_3) = (0, 0.05, 0.2, 0.6)$ was used. Since we used $M = 3$ models, Stage 1 had $\mathcal{X}_1 = 2^M = 8$ possible results and $B = \{\mathbf{b}_1, \dots, \mathbf{b}_8\}$. Our example had $L = 4$ adaptation choices so that the set of adaptations would be $A = \{\mathbf{a}_1, \dots, \mathbf{a}_4\}$. In Stage 2, doses were selected based on ADTAR-rule.

The ADTAR-rule function used the function $g : B \rightarrow A$ to map Stage 1's outcome into dose choices for Stage 2 as follows.

$$g: \begin{cases} \mathbf{b}_1, \mathbf{b}_2 & \rightarrow \mathbf{a}_1 \\ \mathbf{b}_4, \mathbf{b}_6 & \rightarrow \mathbf{a}_2 \\ \mathbf{b}_7 & \rightarrow \mathbf{a}_3 \\ \mathbf{b}_3, \mathbf{b}_5, \mathbf{b}_8 & \rightarrow \mathbf{a}_4 \end{cases}$$

Our example adaptation ADTAR-rule does not allow stopping after Stage 1. This implies that $\alpha_1 = 0$ and $\alpha_0 = 1$ in (5). The contrast weights for ADTAR-rule are computed using (10) in Appendix A and summarized in Table 2. Note that in Table 2 the adaptation choices are numbered from '1' to '4' while they are now formally called $\mathbf{a}_1$ to $\mathbf{a}_4$.

4.2 Evaluation of PoC Probability

We compare probabilities of detecting PoC (PoC probabilities) for our xMCP-Mod method with the four fixed two-stage designs and the MCP-Mod method. Simulation results of the four types of D-R datasets for four total sample sizes (16 scenarios in total) are summarized in Tables 5 to 8 (The R programs for implementing the simulation studies are available upon request).

In this section, 'ADTAR-example' denotes the proposed adaptive two-stage xMCP-Mod design with ADTAR-rule adaptation rule. All four fixed two-stage designs have two components, Study 1 and Study 2. Study 1 of each fixed two-stage design is identical to Stage 1 in ADTAR-example. Study 2 of each fixed two-stage design is identical to one of the adaptive choices of Stage 2 in ADTAR-example. Then, Study 1 and Study 2 are chosen independently of each other. 'Maximum' denotes the fixed two-stage design with the maximum PoC probability and 'Minimum' denotes the fixed two-stage design with the minimum PoC probability, respectively. The one-stage design of the original MCP-Mod method is denoted as 'One-Stage MCP-Mod'.

Data for all four total sample sizes is shown in Tables 5 to 8. Figure 4 plots the global PoC probabilities for total sample size 200 (data taken from Table 6) and compares ADTAR-example, Maximum, Minimum, and One-Stage MCP-Mod for the true D-R curves. We only discuss the detailed results for total sample size 200 since the general trend is similar for all sample sizes.

As shown in Tables 5 to 8, the FWER is controlled roughly at a 5% level in all study designs. The 95% confidence interval for our simulation of 10,000 independent trials with binary outcome (using the binomial distribution) is 500 ± 42.4 trials given a 0.05 true type I error rate. The PoC probability of ADTAR-example is close to the best of all tested study designs. Furthermore, ADTAR-example shows relatively consistent PoC probabilities across the three different D-R curves. ADTAR-example is slightly more powerful than Maximum for the Emax and logistic D-R curves. The difference in PoC probability for the four fixed two-stage designs is considerable (between 5 and 20 percentage points). The PoC probability for One-Stage MCP-Mod is very sensitive to true dose response curve. It shows

higher PoC probability than Maximum for both Emax and Logistic. However, its PoC probability for the quadratic dose-response is lower than Minimum while ADTAR-example shows higher PoC probability than Minimum.

In summary, ADTAR-example is robust and has a comparably high PoC probability. The loss of power due to adaptation is measurable but not high. While the one-stage MCP-Mod can outperform ADTAR-example, its drawback is higher sensitivity than ADTAR-example. One of the four fixed two-stage designs tends to be close to the (or the) best-performing design. However, the actual best fixed two-stage design depends on the underlying true D-R curve, showing their sensitivity.

4.3 Effect of Sample Size Imbalance across Groups

This section evaluates the compensating effect of our weighting procedure for the dependency from data of Stage 1 that produces a bias for the test statistics in Stage 2 as described in Section 3.5. Using our simulation data (four D-R curves for four total sample sizes), we compare ADTAR-example against a variant that does not use weighting in Stage 2's test statistics (denoted as 'Non-weighting ADTAR-example') by computing the respective PoC probabilities. We use aggregate sample size imbalance across multiple trials and PoC probability as measures to evaluate the effect. These PoC results for Non-weighting ADTAR-example are included in Tables 5 to 8 and shown in Figure 4.

First we show that an ADTAR rule can lead to an aggregate sample size imbalance. As shown in Figure 5(a), we consider a hypothetical case where the eight possible PoC results in Stage 1 ($\mathbf{b}_1, \dots, \mathbf{b}_8$) are equally likely to occur in a trial using our two-stage ADTAR design. In this example, if the total sample sizes in both stages are 300, then the aggregate sample sizes for the pre-specified candidate doses, $d_0, \dots, d_4$, are computed as shown in Figure 5(b). The resulting sample sizes are not balanced across doses in 10,000 trials (35% variation). This is because (1) ADTAR-rule (Section 4) does not map the potential results, $\mathbf{b}_1, \dots, \mathbf{b}_8$, to the pre-specified adaptation choices, $\mathbf{a}_1, \dots, \mathbf{a}_4$, in a one-to-one fashion and (2) the number of potential results in Stage 1 that map to the adaptation choices is not the same across the adaptation choices. Furthermore, doses to be used in Stage 2 are not the same across the four adaptation choices. Hence, adjusting the aggregate sample sizes with our proposed weights (Table 2), the sample size imbalance is much improved (5% variation) as shown in Figure 5. This example illustrates how a missing data mechanism in Stage 2 can result using an ADTAR approach and how our weighting based on a missing data analysis can compensate the resulting imbalance.

Second, we prove the effectiveness of the weighting procedure by showing that the compensation of sample sizes improves the PoC probability. As described in Section 3.5, our weighting procedure adjusts for the dependency between Stage 1 data and Stage 2 dose selection via ADTAR that introduces a bias for the test statistics in Stage 2. In simulation studies, we cannot directly measure such a bias. However, when the experimental settings of the weighted and unweighted simulations are the same except for weighting, low PoC probabilities under a true dose-response curve indicate high bias while high PoC probabilities under a true dose-response curve indicate low bias. Tables 5 to 8 show that ADTAR-example outperforms the non-weighting ADTAR-example across the three

different dose-response curves. The gain is up to approximately 3 percentage points, and we show that both weighted and unweighted contrast tests preserve the FWER but the unweighted procedure is less powerful than the weighted version. The results for sample size 200 are included in Figure 4.

4.4 Controlling the FWER

In this section, we evaluate how much our pre-specified ADTAR-rule can control the FWER close to a pre-chosen 0.05 level in a two-stage design. The underlying idea is that fewer choices can lead to a better ability of controlling the FEWR. The ADTAR-rule only uses a subset of all possible adaptation choices (four out of 15 possibilities). To test this idea, we compare the ADTAR-example design with a random two-stage design without ADTAR as follows. We simulated 10,000 trials using a random two-stage design, denoted as ‘Random Two-Stage’, where all 15 dose combinations were equally possible choices in Stage 2 (Figure 2). Table 9 shows both Stage 2’s and global PoC results for four different models (constant, Emax, quadratic, and logistic models) and four different total sample sizes ($N = 80, 200, 400, \text{ and } 600$). When the data has no D-R relationship (under the constant model), Stage 2’s PoC results for the Random Two-Stage design are as low as 0.031 across different total sample sizes. As a result, the global PoC probability under the constant model for the Random Two-Stage designs is also lower than the ideal value of 0.05 (as low as 0.039) for all sample sizes. In contrast, ADTAR-example has a FWER around 0.05 (no lower than 0.048) in this situation.

As we can see in this example, without specifying an ADTAR rule (and thus restricting the number of adaptation choices) in an adaptive two-stage design, the FWER can be over-controlled especially with many dose-combination choices in Stage 2.

5 Concluding Remarks

In certain situations it is necessary to pre-specify doses to use in clinical trials for D-R research. On the other hand, regulatory agencies encourage researchers to develop new adaptive designs for D-R research, and adding and/or dropping doses is one possible adaptation strategy. Enabling prespecification and dose adaptivity at the same time seems contradictory. In this paper we proposed xMCP-Mod method that overcomes this apparent dilemma by providing a framework for dose pre-specification and adaptivity through the pre-specification of an adaptation rule.

Our method is an extension of the MCP-Mod method by Branson et al. and Bretz, et al. from a classical clinical trial design to an adaptive two-stage trial design. Among a set of potential doses, several doses and a placebo were used in the first stage. In the second stage, doses were selected (including the placebo) according to our pre-specified dose adaptation rule called ADTAR (adding and/or dropping treatment adaptation rule). In both stages, we used the same set of the dose-response candidate models. In each stage, the model-associated multiple contrast test statistics were used for the preliminary hypothesis testing on no dose-response relationship. We then combined the test results of both stages in order to establish global PoC via a conditional error function.

We focused on one of the three objectives in the original MCP-Mod method: to establish global PoC after adaptively adding and/or dropping treatment doses in the second stage based on the result of the first stage. In the second stage, we set up an adaptive null hypothesis to resolve all dependencies but sample size imbalance. We showed how to model unequal sample sizes per dose across adaptation choices due to ADTAR function using the missing data weighting method.

Our simulation results showed that our xMCP-Mod method for a given adaptation rule was robust and had a comparably high PoC probability. The original MCP-Mod occasionally outperformed our method, but had the drawback of higher sensitivity to model choices. For completeness, we also compared our method to fixed two-stage designs and showed their higher sensitivity to the true model underlying the data. Showing improved power, we demonstrated the effectiveness of our weighting procedure for Stage 2 in that it compensates for data-dependent variations in sample size per dose. Furthermore, we showed that without specifying an ADTAR rule in an adaptive two-stage design, the FWER could be over-controlled when there were many dose-combination choices in Stage 2.

Our research raises the following questions for further investigation: identifying the best candidate dose-response model and determining a target treatment dose in the proposed design method. Due to logistic considerations, effective sample allocation is another topic of interest in drug development. Hence, in future research we plan to investigate the relationship between sample allocation and PoC probability in the proposed design. Finally, an open question remains about how to find the optimal ADTAR, which depends on previously obtained pre-clinical and clinical data (including toxicity and efficacy) for the respective investigational drug.

Acknowledgments

Stewart J. Anderson was partially supported by NIMH grant: ACISR: P30 MH090333 and NIH/NCI grant U10CA69651. The authors thank the associate editor and two referees for their helpful comments and suggestions.

A Derivation of Weights

In this section, we show the derivation of the weight for adjusting the response estimate of dose group i and adaptation choice j in Stage 2 given an adaptive two-stage design with an ADTAR function as used in Section 3.5. Suppose that we perform a preliminary PoC test using a set of M candidate dose-response models in Stage 1 and Stage 2, respectively. Let B denote the set containing all possible PoC results in Stage 1, i.e., $B = \{\mathbf{b}_1, \dots, \mathbf{b}_{\mathcal{K}_1}\}$ where $\mathcal{K}_1 = 2^M$. Remember that, under the ADTAR, we pre-specify a set of L adaptation choices, $A = \{\mathbf{a}_1, \dots, \mathbf{a}_L\}$ where $\mathbf{a}_j$ denotes choice j . B is partitioned into $B_1 \cup B_2 \cup \dots \cup B_L$ where $B_i \cap B_j = \emptyset$ for $i \neq j$. If we let $|B_j| = \ell_j$ then $\sum_{j=1}^L |B_j| = \sum_{j=1}^L \ell_j = \mathcal{K}_1$. ADTAR is defined by an adaptation function, $g : B \rightarrow A$.

Let $\mathbf{y}_b = (y_{b0}, \dots, y_{bk})$ denote a vector of $k+1$ dose responses in Stage 2 associated with one potential result $\mathbf{b}$, $\mathbf{b} \in B$, in Stage 1. Let $\mathbf{Y} = (y_{bi})$. For result $\mathbf{b}$, we define Q_{bj} as the sample indicator of Stage 2.

$$Q_{bj} = \begin{cases} 1 & \text{if } j = g(\mathbf{b}), \\ 0 & \text{otherwise.} \end{cases}$$

Q_{bj} is used to determine the sample set of Stage 2. Let $\mathbf{Q}_j = (Q_{1j}, \dots, Q_{\mathcal{K}_{1j}})$. The dose adaptation process is characterized by a distribution for $\mathbf{Q}_j$ given both $\mathbf{Y}$ and the adaptive design information, $\mathbf{Z}$. In order to infer quasi-randomization in Stage 2, the following two assumptions are required (Little and Rubin 2002):

1. The distribution of the random indicator variables, $\mathbf{Q}_j$, is determined before any values of $\mathbf{Y}$ in Stage 2 are known. Thus, $f(\mathbf{Q}_j|\mathbf{Y}, \mathbf{Z}) = f(\mathbf{Q}_j|\mathbf{Z})$.
2. Every result $\mathbf{b}$ has a positive (known) probability of being selected. Hence, if $\tau_{bj} = E(Q_{bj}|\mathbf{Y}, \mathbf{Z}) = \Pr(Q_{bj} = 1|\mathbf{Y}, \mathbf{Z})$, then $\tau_{bj} > 0$ is required for all $\mathbf{b}$.

In this context, our adaptation design matrix $\mathbf{Z}$ is a $L \times (k+1)$ matrix indicating the L adaptation ‘strata’ for the $k+1$ doses. Then, both of the assumptions are held and $\tau_{bj} = E(Q_{bj}|\mathbf{Y}, \mathbf{Z}) = E(Q_{bj}|\mathbf{Z}) = E(Q_{bj}|j = g(\mathbf{b})) = 1$. The idea expressed in this equation is as follows: the probability of $j = g(\mathbf{b})$ under the condition that $j = g(\mathbf{b})$ is 1. This statement is the second condition for using the missing data mechanism. Following Little and Rubin, the trial is encoded as $\mathbf{Z}$ and the ADTAR function g is encoded as Q_{bj} . Using $\mathbf{Z}$ and Q_{bj} , our condition can be expressed as $\Pr(Q_{bj} = 1|\mathbf{Y}, \mathbf{Z})$. Thus, $\tau_{bj} = 1$.

Let I_{ij} be an indicator variable and the ji^{th} element of the design matrix, $\mathbf{Z}$, such that $\mathbf{Z}_{L \times (k+1)} = (I_{ij})'$. According to the ADTAR function, I_{ij} is defined as follows.

$$I_{ij} = \begin{cases} 1 & \text{dose group } i \text{ is included in the adaptation choice } j, \\ 0 & \text{otherwise,} \end{cases}$$

where $i = 0, \dots, k$ and $j = 1, \dots, L$. Now, let n_b denote the sample size per dose group in Stage 2 given result $\mathbf{b}$ in Stage 1. Furthermore, let n'_j denote the population size per dose group in Stage 2 in stratum j . If equal sample sizes are assumed in Stage 2 and if the total sample size

is N in Stage 2, then $n'_j = \sum_{g^{-1}(j)} n_b = \ell_j n_b$ and $n_b = \frac{N}{\sum_{i=0}^k I_{ig(b)}}$. If result $\mathbf{b}$ in Stage 1 is associated with adaptation $\mathbf{a}_j$, for each dose group i a stratified random sampling takes a simple random sample scheme of n_b units from n'_j population units within stratum j . In a typical weighting approach, if a unit selected from a target population with probability π represents π^{-1} units in the population, then the unit should be given the weight π^{-1} in estimates of population quantities. In our particular case, π_{ij}^{-1} is used as the weight for the population of dose i in stratum j . In Stage 2, let φ_{ib}^{-1} be the weight for a selected unit of dose i in the j^{th} adaptation based on result $\mathbf{b}$. Since function g maps result $\mathbf{b}$ to stratum j ,

$$\varphi_{ib}^{-1} = \begin{cases} \left(\frac{n_b}{n'_{ij}}\right)^{-1} = \ell_j & \text{if } I_{ij}=1, \\ \text{does not exist} & \text{if } I_{ij}=0, \end{cases} \quad \text{with } j=g(\mathbf{b}).$$

Using ϕ_{ib} , $\pi_{ij} = \phi_{ib} \tau_{bj}$. The Horvitz-Thompson weighted estimator of the population total of dose i , say T_i , over the adaptation choices uses stratum weight, π_{ij} , as follows.

$$t_i HT = \sum_{j=1}^L \pi_{ij}^{-1} y_{ij}. \quad (8)$$

As a result, the stratified mean of dose i is written as:

$$\bar{y}_{i \text{ st}} = \frac{1}{L} \sum_{j=1}^L \phi_{ij} y_{ij}, \quad (9)$$

where $\varphi_{ij} = \frac{L\pi_{ij}^{-1}}{\sum_{v=1}^L \pi_{iv}^{-1}}$. Let $\bar{y}_i^{(w)}$ denote our weighting class estimator of mean response for dose group i , and let $\bar{y}_{ij}$ be the mean response of dose i in adaptation $\mathbf{a}_j$ in Stage 2. Furthermore, let ϕ_{ijb} denote the weight for dose i based on result $\mathbf{b} \in B_j$, then from (9)

$\bar{y}_i^{(w)} = \frac{1}{L} \sum_{j=1}^L \phi_{ijb} \bar{y}_{ij}$. This estimator is further simplified as follows.

$$\begin{aligned} \bar{y}_i^{(w)} &= \frac{1}{L} \sum_{j=1}^L \phi_{ijb} \bar{y}_{ij} \\ &= \frac{1}{L} \sum_{j=1}^L \frac{L(\varphi_{ib} \tau_{bj})^{-1}}{\sum_{v=1}^L (\varphi_{iv} \tau_{bv})^{-1}} \bar{y}_{ij} \\ &= \sum_{j=1}^L \frac{\ell_j I_{ij} \tau_{bj}^{-1}}{\sum_{v=1}^L \ell_v I_{iv} \tau_{bv}^{-1}} \bar{y}_{ij}. \end{aligned}$$

In light of the form of weighted estimator (8), w_{ij} is derived as

$$\begin{aligned} w_{ij} &= \frac{\ell_j I_{ij} \tau_{bj}^{-1}}{\sum_{v=1}^L (\ell_v I_{iv} \tau_{bv}^{-1})} \\ &= \frac{\ell_j I_{ij}}{\sum_{v=1}^L (\ell_v I_{iv})} \end{aligned} \quad (10)$$

As an example, we compute weight $w_{12} = \ell_2 I_{12} / \sum_{v=1}^L (\ell_v I_{1v})$ for dose d_1 in adaptation choice $\mathbf{a}_2$ using (10) and ADTAR-rule in Table 4. First, let $L = 4$ since L is the total number of adaptation choices ($\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$). As $\mathbf{a}_2$ is associated with the subset B_2 in the table, we let $\ell_2 = 2$ which is the number of elements in B_2 ($\mathbf{b}_4$ and $\mathbf{b}_6$). Similarly, we find $\ell_1 = 2$, $\ell_3 = 1$, and $\ell_4 = 4$ from Table 4 so that $w_{12} = 2I_{12} / (2I_{11} + 2I_{12} + I_{13} + 3I_{14})$. The indicator values, I_{11} , I_{12} , I_{13} , and I_{14} in this formula for dose d_1 are the entries of the second column

(associated with d_1) of $\mathbf{Z}_{4,5}$ -matrix shown in Section 3.3: $I_{11} = 0$, $I_{12} = 1$, $I_{13} = 1$, and $I_{14} = 1$. Therefore, $w_{12} = 2I_{12}/(2I_{11} + 2I_{12} + I_{13} + 3I_{14}) = 2 \cdot 1/(2 \cdot 0 + 2 \cdot 1 + 1 + 3 \cdot 1) = 1/3$. This is the weight value for dose d_1 and the second adaptation choice found in the second row and the second column of Table 2.

References

- Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine*. 1999; 18:1833–1848. [PubMed: 10407255]
- Bauer P, Köhne K. Evaluation of experiments with adaptive interim analysis. *Biometrics*. 1994; 50:1029–1041. [PubMed: 7786985]
- Berry, DA.; Müller, P.; Grieve, AP.; Smith, M.; Parke, T.; Balazek, R.; Mitchard, N.; Krams, M. Adaptive Bayesian designs for dose-ranging trials. In: Gatsonis, C.; Kass, RE.; Carlin, B.; Carriquiry, A.; Gelman, A.; Verdine, I.; West, M., editors. *Case Studies in Bayesian Statistics V*. New York: Springer Verlag; 2002.
- Bornkamp B, Bretz F, Dmitrienko A, Enas G, Gaydos B, Hsu C, König F, Krams M, Liu Q, Neuenschwander B, Parke T, Pinheiro J. Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics*. 2007; 17:969–995.
- Bornkamp B, Bretz F, Dette H, Pinheiro J. Response-adaptive dose-finding under model uncertainty. *The Annals of Applied Statistics*. 2011; 5:1611–1631.
- Branson, M.; Pinheiro, J.; Bretz, F. Novartis Biometric Technical Report Number 2003-08-20. 2003. Searching for an adequate dose: combining multiple comparisons and modeling techniques in dose-response studies.
- Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*. 2005; 61:738–748. [PubMed: 16135025]
- Bretz F, Schmidli H, König F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: general concepts. *Biometrical Journal*. 2006; 48(4):623–634. [PubMed: 16972714]
- Chang M. Adaptive design method based on sum of p-values. *Statistics in Medicine*. 2007; 26:2772–2784. [PubMed: 17133651]
- Derendorf H, Meibohm B. Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: concepts and perspectives. *Pharmaceutical Research*. 1999; 16(2):10.
- Dragalin V, Fedorov V. Adaptive designs for dose-finding based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*. 2006; 136:1800–1823.
- Dragalin V, Bornkamp B, Bretz F, Miller F, Padmanabhan SK, Patel N, Perevozskaya I, Pinheiro J, Smith JR. A simulation study to compare new adaptive dose-ranging designs. *Statistics in Biopharmaceutical Research*. 2010; 2(4):487–512.
- FDA. Guidance for industry: Adaptive design clinical trials for drugs and biologics (Draft guideline). 2010. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf> (Accessed on 17 July, 2010).
- Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J. Adaptive designs in clinical drug development – an executive summary of the PhRMA working group. *Journal of Biopharmaceutical Statistics*. 2006; 16:275–283. [PubMed: 16724485]
- Gaydos B, Dragalin V, Coffey C. Fall 2007 web-based training series. Webinar 2. Adaptive designs in dose-finding studies. *Biopharmaceutical Section of ASA*. 2007
- Hiyama Y. About quality Control for test articles in exploratory clinical studies. *Health and Labour Sciences Special Research Report*, in Japanese. 2007
- Hommel G. Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal*. 2001; 5:581–589.
- Hung HMJ, O'Neill RT, Wang SJ, Lawrence J. A regulatory view on adaptive/flexible clinical trial design. *Biometrical Journal*. 2006; 48(4):565–573. [PubMed: 16972707]

- ICH. ICH Harmonized Tripartite Guideline E4: Dose-response information to support drug registration. 1994. Available at: <http://www.ich.org/LOB/media/MEDIA480.pdf> (Accessed on 17 July, 2010).
- Little, RJA.; Rubin, DB. Wiley Series in Probability and Statistics. 2. 2002. Statistical analysis with missing data.
- Meibohm B, Derendorf D. Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modeling. International Journal of Clinical Pharmacology and Therapeutics. 1997; 35(10):401–413. [PubMed: 9352388]
- Miller F. Adaptive dose-finding: Proof of concept with type I error control. Biometrical Journal. 2010; 52(5):577–589. [PubMed: 20976691]
- Phillips AJ, Keene ON. Adaptive designs for pivotal trials: discussion points from the PSI adaptive design expert group. Pharmaceutical Statistics. 2006; 5:61–66. [PubMed: 17080929]
- Pinheiro J, Bornkamp B, Bretz F. Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. Journal of Biopharmaceutical Statistics. 2006; 16:639–656. [PubMed: 17037263]
- Pinheiro J, Sax F, Antonijevic Z, Bornkamp B, Bretz F, Chuang-Stein C, Dragalin V, Fardipour P, Gallo P, Gillespie W, Hsu C, Miller F, Padmanabhan SK, Patel N, Perevozskaya I, Roy A, Sanil A, Smith JR. Adaptive and model-based dose-ranging trials: quantitative evaluation and recommendations. White Paper of the PhRMA working group on adaptive dose-ranging studies. Statistics in Biopharmaceutical Research. 2010; 2(4):435–454.

Figure 1.

Evaluation: Adaptive Design with ADTAR-rule, Fixed Two-stage Designs, and One-stage MCP-Mod Design.

(a). Adaptive Design with ADTAR-rule.

(b). Fixed Two-stage Designs (Four Study Combinations).

(c). One-stage MCP-Mod Design.

Figure 2.
Random Two-Stage design.

(a)

(b)

(c)

Figure 3.
 Standardized Emax (f_1), quadratic (f_2), and logistic (f_3) models.
 (a). Standardized Emax model: f_1 .
 (b). Standardized quadratic model: f_2 .
 (c). Standardized logistic model: f_3 .

Figure 4. PoC Probabilities in Global Stage for Different Dose-Response Patterns in ADTAR-example, Non-weighting ADTAR-example, One-Stage MCP-Mod, Maximum, and Minimum designs.

(a)

(b)

(c)

Figure 5.
 Equal probability of Stage 1's results and sample sizes with and without weighting.
 (a). Stage 1's potential results.
 (b). Sample sizes without weighting.
 (c). Sample sizes with weighting.

Table 1

xMCP-Mod Adaptation Rule for Doses Used in Stage 2, Called ADTAR-rule

Model significance		Doses in Stage 2						
f_1	f_2	f_3	d_0	d_1	d_2	d_3	d_4	
1	1	1		1	0	1	1	1
1	1	0						
0	1	1		1	1	1	0	1
0	1	0						
0	0	1	1	1	1	1	1	0
1	0	1						
1	0	0	1	1	1	1	1	1
0	0	0						

For model significance, 1 indicates significant model in Stage 1 and 0 otherwise. For doses in Stage 2, 1 indicates doses to be used in Stage 2 and 0 otherwise. In this example, there are eight possible inputs, which can be enumerated as the number of rows in the left column, for the decision making on dose selection. There are 4 rows in the right column which is the number of pre-specified dose combinations called adaptation choices 1, 2, 3, and 4.

Table 2

Weights Under The Adaptation Choices in Table 1

	1	2	3	4
d_0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{8}$
d_1	0	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$
d_2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{8}$
d_3	$\frac{1}{3}$	0	$\frac{1}{6}$	$\frac{1}{2}$
d_4	$\frac{2}{7}$	$\frac{2}{7}$	0	$\frac{3}{7}$

Table 3

Original and Standardized Model Formula

Model	Original formula	Standardized formula
f_1 : Emax	$f_1(d) = E_0 + \frac{E_{\max} d}{E D_{50} + d}$	$f_1^0(d) = \frac{d}{E D_{50} + d}$
f_2 : Quadratic	$f_2(d) = E_0 + \beta_1 d + \beta_2 d^2$	$f_2^0(d) = d + \frac{\beta_2}{ \beta_1 } d^2$ $= d - \delta_{\max} d^2$
f_3 : Logistic	$f_3(d) = E_0 + \frac{E_{\max}}{1 + \exp\left(\frac{E D_{50} - d}{\delta}\right)}$	$f_3^0(d) = \frac{1}{1 + \exp\left(\frac{E D_{50} - d}{\delta}\right)}$
$d \in \{d_0, d_1, \dots, d_4\}$		

Table 4

ADTAR-rule

B_j	b	Significant Models in Stage 1	I_j	a_j	doses to add	doses to drop	doses used in Stage 2	balanced sample, n_{2i} for $n_{1i} = 25$ (n_{2i} for $n_{1i} = 75$)
B_1	b_1	f_1, f_2, f_3	2	a_1	d_4	d_1	d_0, d_2, d_3, d_4	25 (75)
	b_2	f_1, f_2						
B_2	b_4	f_2, f_3	2	a_2	d_4	d_3	d_0, d_1, d_2, d_4	25 (75)
	b_6	f_2						
B_3	b_7	f_3	1	a_3	None	None	d_0, d_1, d_2, d_3	25 (75)
B_4	b_3	f_1, f_3	3	a_4	d_4	None	d_0, d_1, d_2, d_3, d_4	20 (60)
	b_5	f_1						
	b_8	None						

Table 5

PoC probabilities of ADTAR-example, Non-weighting ADTAR-example, One-Stage MCP-Mod, and the four two-stage fixed designs under varying D-R shape for total sample size 80

Design Choice	Test Stage	n_1/gp n_2/gp (N)	Data-Generating Function			
			Constant	E _{max}	Quadratic	Logistic
ADTAR-example	Stage 1	10	0.0369	0.1679	0.2070	0.1811
	Stage 2	8–10	0.0467	0.2243	0.1648	0.2806
	Global	(80)	0.0499	0.3206	0.3039	0.3791
Non-weighting ADTAR-example	Stage 2	8–10	0.0450	0.2038	0.1428	0.2550
	Global	(80)	0.0490	0.3089	0.2884	0.3643
One-Stage MCP-Mod	Global	(80)	0.0424	0.3294	0.2262	0.4286
a_I	Stage 1	10	0.0397	0.1759	0.2130	0.1888
	Stage 2	10	0.0455	0.2210	0.1520	0.2650
	Global	(80)	0.0503	0.3329	0.2999	0.3804
a_{II}	Stage 1	10	0.0381	0.1679	0.2103	0.1792
	Stage 2	10	0.0390	0.2038	0.1012	0.2264
	Global	(80)	0.0440	0.3044	0.2470	0.3439
a_{III}	Stage 1	10	0.0361	0.1704	0.2032	0.1842
	Stage 2	10	0.0394	0.1716	0.2068	0.1781
	Global	(80)	0.0447	0.2747	0.3515	0.3043
a_{IV}	Stage 1	10	0.0345	0.1688	0.2145	0.1837
	Stage 2	8	0.0439	0.2078	0.1503	0.2590
	Global	(80)	0.0501	0.3146	0.2992	0.3723

Table 6

PoC probabilities of ADTAR-example, Non-weighting ADTAR-example, One-Stage MCP-Mod, and the four two-stage fixed designs under varying D-R shape for total sample size 200

Design Choice	Test Stage	n_1/gp n_2/gp (N)	Data-Generating Function			
			Constant	E _{max}	Quadratic	Logistic
ADTAR-example	Stage 1	25	0.0382	0.3207	0.4100	0.3559
	Stage 2	20-25	0.0467	0.4408	0.3266	0.5544
	Global	(200)	0.0483	0.6102	0.5896	0.7134
Non-weighting ADTAR-example	Stage 2	20-25	0.0422	0.3908	0.2679	0.4928
	Global	(200)	0.0461	0.5888	0.5570	0.6844
	Global	(200)	0.0418	0.6428	0.4484	0.7739
a_I	Stage 1	25	0.0369	0.3299	0.4120	0.3620
	Stage 2	25	0.0420	0.4171	0.2668	0.5111
	Global	(200)	0.0484	0.6017	0.5614	0.6952
a_{II}	Stage 1	25	0.0368	0.3277	0.4093	0.3528
	Stage 2	25	0.0366	0.3816	0.1547	0.4462
	Global	(200)	0.0426	0.5907	0.4626	0.6556
a_{III}	Stage 1	25	0.0372	0.3241	0.4067	0.3551
	Stage 2	25	0.0410	0.3289	0.4118	0.4075
	Global	(200)	0.0444	0.5472	0.6625	0.6238
a_{IV}	Stage 1	25	0.0390	0.3237	0.4213	0.3542
	Stage 2	20	0.0437	0.4028	0.2702	0.4982
	Global	(200)	0.0466	0.5951	0.5675	0.6888

Table 7

PoC probabilities of ADTAR-example, Non-weighting ADTAR-example, One-Stage MCP-Mod, and the four two-stage fixed designs under varying D-R shape for total sample size 400

Design Choice	Test Stage	n_1/gp n_2/gp (N)	Data-Generating Function			
			Constant	E _{max}	Quadratic	Logistic
ADTAR-example	Stage 1	50	0.0403	0.5411	0.6667	0.5979
	Stage 2	40-50	0.0464	0.7327	0.5658	0.8261
	Global	(400)	0.0504	0.8747	0.8555	0.9369
Non-weighting ADTAR-example	Stage 2	40-50	0.0417	0.6487	0.4409	0.7587
	Global	(400)	0.0478	0.8551	0.8261	0.9205
One-Stage MCP-Mod	Global	(400)	0.0411	0.8879	0.7032	0.9618
a_I	Stage 1	50	0.0386	0.5378	0.6685	0.5878
	Stage 2	50	0.0454	0.6569	0.4245	0.7817
	Global	(400)	0.0481	0.8645	0.8246	0.9259
a_{II}	Stage 1	50	0.0368	0.5422	0.6636	0.5837
	Stage 2	50	0.0405	0.6306	0.2462	0.7159
	Global	(400)	0.0477	0.8493	0.7368	0.8999
a_{III}	Stage 1	50	0.0396	0.5514	0.6693	0.5883
	Stage 2	50	0.0379	0.5467	0.6664	0.5888
	Global	(400)	0.0446	0.8170	0.9070	0.8521
a_{IV}	Stage 1	50	0.0421	0.5469	0.6673	0.5853
	Stage 2	40	0.0448	0.6448	0.4411	0.7750
	Global	(400)	0.0505	0.8576	0.8268	0.9247

Table 8

PoC probabilities of ADTAR-example, Non-weighting ADTAR-example, One-Stage MCP-Mod, and the four two-stage fixed designs under varying D-R shape for total sample size 600

Design Choice	Test Stage	n_1/gp n_2/gp (N)	Data-Generating Function			
			Constant	Emax	Quadratic	Logistic
ADTAR-example	Stage 1	75	0.0393	0.7054	0.8205	0.7601
	Stage 2	60-75	0.0468	0.8865	0.7340	0.9434
	Global	(600)	0.0521	0.9674	0.9527	0.9893
Non-weighting ADTAR-example	Stage 2	60-75	0.0397	0.8008	0.5722	0.9005
	Global	(600)	0.0473	0.9570	0.9382	0.9845
One-Stage MCP-Mod	Global	(600)	0.0446	0.9716	0.8663	0.9949
a_I	Stage 1	75	0.0422	0.7053	0.8287	0.7508
	Stage 2	75	0.0447	0.8177	0.5752	0.9117
	Global	(600)	0.0514	0.9601	0.9409	0.9844
a_{II}	Stage 1	75	0.0378	0.7117	0.8231	0.7670
	Stage 2	75	0.0401	0.7863	0.3355	0.8582
	Global	(600)	0.0450	0.9498	0.8799	0.9772
a_{III}	Stage 1	75	0.0370	0.7098	0.8297	0.7594
	Stage 2	75	0.0351	0.7074	0.8322	0.7598
	Global	(600)	0.0405	0.9318	0.9793	0.9541
a_{IV}	Stage 1	75	0.0364	0.6961	0.8273	0.7561
	Stage 2	60	0.0446	0.7991	0.5857	0.8977
	Global	(600)	0.0486	0.9527	0.9397	0.9831

Table 9
PoC probabilities of Random Two-Stage design under varying D-R shape $\times$ sample size

Design Choice	Test Stage	n_2/gp (N)	Data-Generating Function			
			Constant	E _{max}	Quadratic	Logistic
Random Two-Stage	Stage 2	8–20	0.0354	0.1874	0.1337	0.2012
	Global	(80)	0.0438	0.2975	0.2845	0.3098
Random Two-Stage	Stage 2	20–50	0.0319	0.3688	0.2673	0.3970
	Global	(200)	0.0391	0.5705	0.5544	0.5937
Random Two-Stage	Stage 2	40–100	0.0372	0.5896	0.4242	0.6060
	Global	(400)	0.0437	0.8347	0.8113	0.8318
Random Two-Stage	Stage 2	60–150	0.0359	0.7377	0.5533	0.7090
	Global	(600)	0.0435	0.9295	0.9261	0.9145