OFFICE OF CLINICAL PHARMACOLOGY
DIVISION OF PHARMACOMETRICS

Application	Request for Qualification of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty
Applicant	Janssen Pharmaceuticals and Novartis Pharmaceuticals
Application date	22 April, 2015
OCP Division	Division of Pharmacometrics
OCP Reviewer	Dinko Rekić, MSc(Pharm), Ph.D.
Concurring reviewers	Yaning Wang, Ph.D. Deputy Director, Division of Pharmacometrics Vikram Sinha, Ph.D. Director, Division of Pharmacometrics

OCP: Office of Clinical Pharmacology

Table of Contents

1 EXECUTIVE SUMMARY

Janssen Pharmaceuticals and Novartis Pharmaceuticals have submitted a "Request for Qualification of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty". MCP-Mod, (Multiple Comparison Procedure – Modelling) is a hybrid approach that combines hypothesis testing and modeling in a structured manner to analyze phase 2 dose-ranging studies with the purpose of finding suitable dose(s) for confirmatory phase 3 trials.

The first step of the procedure (MCP-step) is used to assess presence of a dose-response signal using a trend test deducted from a set of prespecified candidate models. The second step (Mod-step) relies on parametric modeling or model averaging to find the "optimal' dose for confirmatory trials. MCP-Mod method is found more effective than pairwise comparison due to its ability to utilize all available data from the continuum of active doses and placebo to estimate a parametric dose-response curve which allows for interpolation and extrapolation of effect across a range of doses. Other structured hybrid testing-modeling approaches are likely to provide similar benefit. Similar to other methods that depend on modeling, MCP-Mod encourages better design of dose-ranging studies by allocation of doses at the ascending part of the dose-response curve.

1.1 Recommendations

The Division of Pharmacometrics in The Office of Clinical Pharmacology finds the MCP-Mod method an adequate and appropriate method for dose selection based on dose-response data.

2 SUMMARY OF METHODOLOGY ASSESSMENT

The Applicant's motivation for developing a hybrid testing-modeling approach was to combine the strengths of hypothesis testing (few assumptions, control for type I error) and modeling (better predictive performance) to support better decisions.

This application is supported by: a) summaries of two simulation studies that compared MCP-Mod to other methods, b) examples of use of MCP-Mod in literature, and c) examples of use at the two firms supporting this application.

The focus of this review has been on the evaluation of MCP-mod in the two simulation studies where findings from the applicant's simulations have been confirmed by the reviewer in an independent analysis.

The two simulation studies evaluated MCP-Mod, based on power to detect a significant dose-response relationship, type I error rate, and likelihood to find a minimal effective dose. Methods were evaluated under various model shapes, sample size, and number of doses.

Power to declare a positive dose-response relationship with the MCP-Mod method was compared to Dunnett's test (Dunnett 1955), Log-likelihood ratio (LLR) test for monotonicity (Bartholomew 1961), a Bayesian modeling approach, and a nonparametric modeling approach. Dunnett's test was consistently inferior to other methods. All other methods were similar in performance.

It is noteworthy that the power of the MCP-Mod method was shown to be sensitive to the initial 'guesses' of model shapes. Under a scenario where the simulated model shape was not included amongst the prespecified model shapes, power was 14%-points lower than the average power for the other model shapes. It is therefore advisable not to restrict the number of shapes to be included in the procedure. Power of MCP-Mod method was also negatively affected when number of doses increased but total sample size remained the same, but less so than Dunnett's test. A plausible explanation is that the design with fewer doses included more subjects at highly informative doses compared to the design with a large number of doses. These findings were confirmed in an independent analysis conducted by the reviewer which showed that MCP-Mod performance increased if doses were chosen based on the D-optimal design compared to other designs that included more dose levels. It should be noted that both the model shape and model parameters need to be known *a priori* in order to design a dose-ranging study based on the D-optimal design. This is almost never the case in clinical drug development. It is therefore good practice to explore a wide range of doses so that the shape of the dose response curve can be characterized.

Reviewer's analysis also shows that MCP-Mod performance is similar to LLR test unless the assumption of monotonicity is violated. In that case, MCP-Mod was superior to LLR test.

Applicant's simulations as well as reviewer's analysis show that MCP-Mod method has similar power and type I error to test for significant dose-response relationships as other methods.

MCP-Mod performance was similar to other modeling methods or hybrid methods in terms of power to detect the minimal effective dose.

Applicant's hybrid approach to dose finding is not unique. Similar hybrid methods have been proposed (Overgaard et al. 2015, and Aoki et al. 2014). There is currently no direct comparison between the method proposed by Aoki et al. and the MCP-Mod method however, based on the reviewer's cross-study comparison, there is no reason to believe that there would be a significant difference in performance.

3 INTRODUCTION

Janssen Pharmaceuticals and Novartis Pharmaceuticals have submitted a "Request for Qualification of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty" to the Agency on April 22, 2015. The applicant refers to past interactions with the Agency where the MCP-Mod method was discussed: 1) 1-day course to the CDER/OTS Office of Biostatistics on "Dose Finding Studies: Methods and Implementation" on August 12, 2008. 2) Half-day tutorial on "MCP-Mod: A structured approach to model-based design and analysis of Phase II dose finding studies under model uncertainty" presented to CDER/OTS/OB and OCP on January 28, 2015.

The European Medicines Agency (EMA) has offered a positive qualification opinion on MCP-Mod procedure (EMA 2014).

4   BACKGROUND

Dose selection is a critical component in drug development. Uncertainties regarding the dose can lead to delayed or denied NDA/BLA applications or post-marketing dose changes (Sacks et al. 2014; Cross et al. 2002). According to the ICH-E4 guidance, dose selection for confirmatory trials should be based on the observed dose (concentration) response relationship obtained from earlier studies (phase 1 and phase 2), (ICH-E4 1995). The guidance also emphasizes that dose-response studies should focus on elucidation of the dose-response function, not individual pairwise comparisons between dose groups. Several study designs are proposed e.g. parallel, cross-over, forced titration, optional titration. A minimum of 2 doses in addition to placebo is proposed but more than the minimum is desired. Spacing of doses is recommended to be informed by the between subject pharmacokinetic variability of the drug to avoid extensive exposure overlap between doses.

Based on a meta-analysis of phase 2 studies with non-oncology small molecules approved between 2009 and 2014, most sponsors investigate 2 or more active doses in Phase 2 trials, **Table 1** (Thomas [a] et al. 2014).

Table 1:   **Number of Doses Tested in Phase 2 for FDA Approved Drugs**

Number of dose groups in Phase 2 trials (including placebo)	FDA approved drugs[1] N=65 n (%)	
<3	19	(29.2)
4	23	(35.4)
5	10	(15.4)
6	11	(16.9)
7	2	(3.1)

[1] *Non-oncology small molecules approved between 2009-2014*
Adapted from Thomas [a] et al. 2014

Two guiding principles focusing on either *estimation* or *testing*, have dominated the debate on how dose-response trials should be analyzed. The current application attempts to unify the two principles under a single approach referred to as multiple contrasting and modeling procedure (MCP-Mod).

Testing refers to multiple comparisons in order to formality reject a null hypothesis while estimation refers to estimating the treatment effect for a continuum of doses or concentrations.

The benefits and caveats of each method are discussed in sections 4.1 and 4.2. An overview of applicant's methodology is presented in section 5.1. Results from reviewer's simulation study are presented in section 6.

4.1 Estimation Methods

The use of estimation methods[1] (as opposed to testing methods) to analyze dose-response data is supported in the ICH-E4 guidance and in literature. These recommendations are in line with the learning rather than the confirming purpose of phase 2 studies. Under the 'learning and confirming' paradigm; phase 2b constitutes a 'learning' phase in drug development where the study design and the analysis are geared to maximize information gained from the trial (Sheiner 1997).

Estimation based analysis utilizes observations from the continuum of active doses and placebo to estimate the dose-response curve. Once the dose-response curve is estimated, effect size between tested doses (interpolation) as well as effect size beyond the range of the tested doses (extrapolation) can be estimated with various degrees of precision that depend on the position of the observed doses on the dose-response curve and the underlying assumption about of the relationship between dose and response. Assumption regarding the relationship between dose and response takes form of a parametric relationship (a model) that is driven by the observed data. It is generally accepted that estimation driven or model-based analysis will lead to more precise and less biased estimation of the dose-response relationship due to utilization of all information across doses.

In principal, a dose-response relationship is selected based on how well it describes the observed data. Physiologically plausible dose-response relationships are evaluated based on various goodness of fit criteria in a process referred to as model selection. Following model selection, a single model is selected as the 'best' representation of observed data and subsequently used to estimate treatment effect. Estimates of treatment effect often account for uncertainty in parameter estimates giving a probability range of the typical response along the continuum of doses. While parameter uncertainty is a function of the information content of the observed data, it does not account for uncertainty in model selection. Uncertainty in model selection can be accounted for by averaging estimates of treatment effect over several or all tested models.

It has been argued that ignoring uncertainty in model selection can lead to over-confidence in decision-making (Chatfield 1995)[2]. It should be noted that Chatfield makes this argument for any decision-making that is based on modeling. As such, there can be an infinitive number of models. Given the finite number of shapes a dose-response curve can take, it is debatable whether or not that argument is applicable to dose-response modeling. A recent meta-analysis of a drug portfolio consisting of 33 drugs and 76 dose-ranging studies found that the maximum response (E_{max}) model adequately described all but one drug. In fair-

---

[1] Estimation methods are often referred to as modeling methods where the aim is to describe the relationship between two observed variables with mathematical equations.

[2] This argument has been brought forward by the applicant as to why modeling approaches alone may lead to over confident decisions. Request for FDA Qualification Opinion section 2.3

ness, it should be noted that in this meta-analysis no dose-ranging study included less than 3 active doses (Thomas [b] et al. 2014). Thus, most dose-response relationships are adequately described by an E_{max} model and the impact on model uncertainty on dose selection as it pertains to the shape of the curve may be low.

Given that the relationship between dose and response is driven by the observed data, it is sometimes argued that analysis based estimation is a post-hoc analysis that has limited value. Prespecification of dose-response analysis in combination with model averaging have been proposed to address the issues with estimation based dose-response analysis (Aoki et al. 2014). Others argue that because phase 2 is geared toward learning (hypothesis generation); there is no need to fully pre-specify the analysis and a reasonably predefined and standardized data driven model-finding process is appropriate (Sheiner 1997).

One can argue that estimation methods do not formally test a hypothesis and that there is a risk of selecting a model that describes a significant dose-response relationship when there is none. A common method to guard against that type of type I error is to first formally test the dose-response relationship with a linear model. This method is described in more detail in section 4.2 (Likelihood Ratio Test Statistic for Monotonicity) and is recommended in proposed good practice principals published in literature (Overgaard et al. 2015). Alternatively, as proposed by Aoki and colleagues, one can use the log-likelihood ratio test to compare each drug-effect model to a no-effect model in order to test for significant dose-effect relationship. This method has many similarities with the MCP-Mod procedure, for example: both methods combine hypothesis testing to guard against type I error (log likelihood test and multiple contrast procedure) with estimation methods to select the minimal therapeutic dose. A direct comparison against these methods is currently not available.

4.1.1   The Utility of Concentration-Response Modeling

The MCP-Mod method focuses on analysis of dose-response; however, exposure-response analysis is a fundamental part of dose selection, as such it needs to be put in context of this application.

Dose-response analysis assumes that all subjects in one dose group are exposed to the same amount of drug. Due to between subject variability in drug absorption, distribution, metabolism, and elimination (ADME), subjects administered the same dose may have vastly different drug concentrations in blood. In addition to variability in drug exposure, there is variability in drug response.

Thus i, if we can identify what concentration results in a particular effect we can answer the question what dose will result in a particular concentration.

Exposure-response modeling can also inform the need for dose adjustment in specific populations that may have different drug exposure compared to the typical patient, e.g. due to renal or hepatic impairment. Without relating exposure to response, there is no way of determining the need to adjust dose due to difference in drug exposure.

The main caveat of exposure-response modeling is that exposure in a subject is not randomized (some randomization is present because different doses result in different exposure ranges). Since exposures are not truly random, there is a possibility that a patient factor affecting drug exposure is also affecting drug response (independent of drug exposure). This phenomenon is commonly observed for large molecules with oncology indications where tumor burden can increase the elimination of drugs; however, this phenomenon is rare for small molecules (Yang et al. 2013.)

Another assumption about exposure-response modeling is that the treatment effect is related to drug concentrations in blood. This is known not to be true for inhaled COPD and Asthma drugs where the treatment effect is mediated by local drug concentrations at the site of administration. In such cases, dose response analysis is more informative than exposure-response analyses. However, drug safety, an important element of dose selection is related to drug concentrations in blood is well suited for exposure response analysis, e.g. endogenous cortisol suppression by inhaled corticosteroids for treatment of Asthma.

4.2 Testing Methods

The common core of most testing methods is the treatment of dose as a categorical factor (as opposed to a continuum) and that few (if any) assumptions are made concerning the underlying dose-response relationship. Treating a continuous variable, such as dose, as a categorical variable leads to loss of statistical as well as predictive power. Statistical power refers to decreased likelihood of detecting a true relationship. Loss of predictive power refers to limiting the inferences one can make based on the analysis e.g. diminished ability to estimate effect between or beyond the observed doses.

Estimation of overall dose-response is commonly based on analysis of variance (ANOVA), with dose as a factor. ANOVA methods can be extended to include other predictors such as patient's baseline characteristics, or study design elements such as period and sequence.

Several multiple comparison procedures are available to differentiate the effect between doses or treatment arms. These include Williams' procedure (Williams, 1971, 1972) where active dose are compared versus placebo in a stepwise manner (starting from the highest). The procedure tests each dose level but stops at the dose level where the null hypothesis (of no dose effect) is not rejected. Williams' procedure assumes that effect increases with increasing dose. A popular alternative to Williams' procedure is Dunnett's test, also referred to many-to-one comparison. Dunnett's test reduces the number of comparison by only testing active doses versus placebo (Dunnett et al. 1955). Another method is the Likelihood Ratio Test Statistic for Monotonicity. The global likelihood ratio test (Bartholomew 1961) can be used to test the null hypothesis of no treatment effect against the alternative hypothesis of where effect increases with dose. The mean effect for each treatment group or dose is estimated with linear regression with or without a slope parameter relating dose to effect. The log likelihood (LL) under the null or the alternative model is compared. When the LL ratio of the two LL is large the null hypothesis is rejected. The threshold for rejection depends on the desired level of significance. A major drawback of this test is the assumption of increasing effect with increasing dose, as is the case for Williams' procedure. A violation of that assumption leads to dramatic decrease in power of the test, as shown in this review and elsewhere.

4.3 Determination of the Minimal Effective Dose

Drug safety data obtained from phase 1 and 2 studies may not always be predictive of the drug safety profile at long-term drug exposure observed in large phase 3 trials. It is therefore often advisable to select a lowest dose that will provide a desired therapeutic effect, i.e. the minimal effective dose (MED). It is important to emphasize that there are no statutory requirements to find the minimal effective dose. It is however, an important inference that can be gleaned from a well conducted dose-finding study. Finding the MED can minimize unnecessary drug exposure that will not lead to additional benefit to the patient but may increase the risk or severity of adverse events.

Several estimation (modeling) methods have been proposed to find the MED. Early examples include the work of Filloon, who was an advocate of estimating the smallest dose with therapeutic benefit as opposed to the smallest dose that is statistically significant from placebo, (Filloon 1995). He argued that since statistical significance is a function of response variability and sample size, larger samples size and smaller response variability would result in lower MED, completely disconnecting MED selection from desired therapeutic effect. Fillon also recognized the problem with estimating the typical treatment effect at a given dose as opposed to estimating the percent of subjects that will reach the desired treatment effect. He advocated for the use of estimation methods to characterize the dose-response curve for the typical patient and to estimate percentiles for each estimated dose effect in order to select a dose that would benefit a predefined number of subject e.g. 75% of the tested population.

Since the works of Filloon, others have contributed to estimation based methods for determining the MED and this practice is now well established in clinical drug development. Recent focus has been on estimating the probability of the typical response to reach a predefined threshold of clinical relevant effect. The estimated probability can be adjusted for multiple levels of uncertainty e.g. uncertainty in parameter estimates and uncertainty in model selection, etc. (Lalonde et al. 2007, Aoki et al. 2014). As mentioned previously, the MCP-Mod procedure relies on modeling methods to estimate the minimal effective dose; an example of this is shown in section 5.2.1.

Several methods to identify the minimal effective dose by testing methods have been proposed. Both the Dunnett's and Williams test have been developed for that purpose. In addition to those methods, step contrasts have also been widely used for finding the minimal effective dose. Bauer and Hackl have shown contrast to be superior compared to Dunnett's and Williams' test, (Bauer and Hackl, 1985).

Applicant has submitted evaluations of different methods to estimate the MED, see section 5.2.

5   APPLICANT'S SUBMISSION

The documentation submitted by the applicant in support of the Agencies qualification opinion is shown in **Table 2.** The submission does not contain original datasets or program codes. The MCP-Mod method is implemented in the publicly available software for statistical computing and graphics: R^3. The review focused on the Applicant's simulation studies that compared the MCP-Mod method to other methods.

Table 2:   Applicant's submitted documentation in support of qualification opinion

Document	Content
Request for Qualification Opinion to the Agency	Methodological aspects of MCP-Mod Examples of use of MCP-Mod in literature and at Novartis and Jensen Summary of simulation studies comparing MCP-Mod to other methods*
Request for Qualification Opinion to the EMA	Similar content as above
Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty	CHMP qualification opinion

Note: * Indicates the focus of this review.

5.1   Overview of the MCP-Mod Procedure

The MCP-Mod procedure consists of a study design step and a data analysis stage. In the study design stage, a set of candidate models are selected based on the known pharmacology of the drug and any relevant previous experience with the indication in question.

There are two analytical steps within the MCP-Mod procedure, the Multiple contrast step (MCP) and the Modeling (Mod) step. The Multiple Contrast step aims to identify a statistically significant dose-response signal. A significant dose-response signal is declared if one or more prespecified models have been found to be significant. The statistically significant models form a model pool used to estimate the dose-response relationship and select a dose for further testing (Modeling step).

---

[3] https://cran.r-project.org/web/packages/DoseFinding/index.html

5.2   Applicant's Comparison of MCP-Mod to Other Methods

The applicant has submitted results from 2 simulation studies. The purpose of these studies was to compare MCP-Mod to other established methods for testing a positive dose-response relationship as well as finding the MED. Applicant's results are summarized below. Reviewer's own interpretation and conclusions are reported in the main text and separately after each section.

5.2.1   MCP-Mod versus LLR test

Simulation Study 1

The first simulation study compared MCP-Mod to the LLR test and the contrast test (Bauer and Hackl, 1985). This study was published by Branson et al. in 2003.

Several dose-response shapes were used to simulate parallel design trials with 5 dose groups (including placebo), **Figure 1**. Equal sample size per dose group was assumed and ranged from 10 to 150 subjects per group. The null hypothesis of a non-significant dose-response relationship was tested at one sided alpha of 0.05. The standard deviation was set at 1.478. A power of 80 % is achieved at a sample size of 75 for the pairwise test at the maximum effect 0.6. A total of 10.000 simulations were performed under each shape and sample size.

Results of simulation study 1 are shown in **Table 3.** Only results from sample size of n=75 are shown.

The following inferences are made with regards to power to detect a true dose-response relationship:[4]

- The MCP-Mod method was numerically superior to LLR for all model shapes except for convex shape.
- The MCP-Mod method was numerically superior to Step Contrast for all model shapes except for convex shape.
- The median power difference between MCP-Mod and LLR was 1.45%, in favor of MCP-Mod.
- The median difference between MCP-Mod and step contrast was 2.85%, in favor of MCP-Mod.
- The median difference between LLR and step contrast was 0.8%, in favor of LLR.

The following inferences are made with regards to type I error rate.

- All methods performed equally and adequately in terms of type I error rate.

The reviewer as well as the applicant agree that power is very similar between LLR and MCP-Mod. MCP-Mod was inferior to LLR for the convex shape, likely

---

[4] Power is defined as percent of the simulated trials where a significant dose-response relationship was found.

because that model was not included in the MCP-Mod library. Power to detect a true significant dose-response under the convex shape was 14% lower compared to the average power of the other shapes. This reviewer thinks that this is an indication of how sensitive the performance of MCP-Mod is to the prespecification of the model library. This is in contrast to the robustness of LLR under the monotone dose-response shape assumption. In cases where the assumption of monotonicity is violated (quadratic and the double-logistic shapes), LLR test performance is on average 16% lower compared to shapes that are monotone.

Figure 1: Dose-response Models Included in Applicants Simulation Study 1
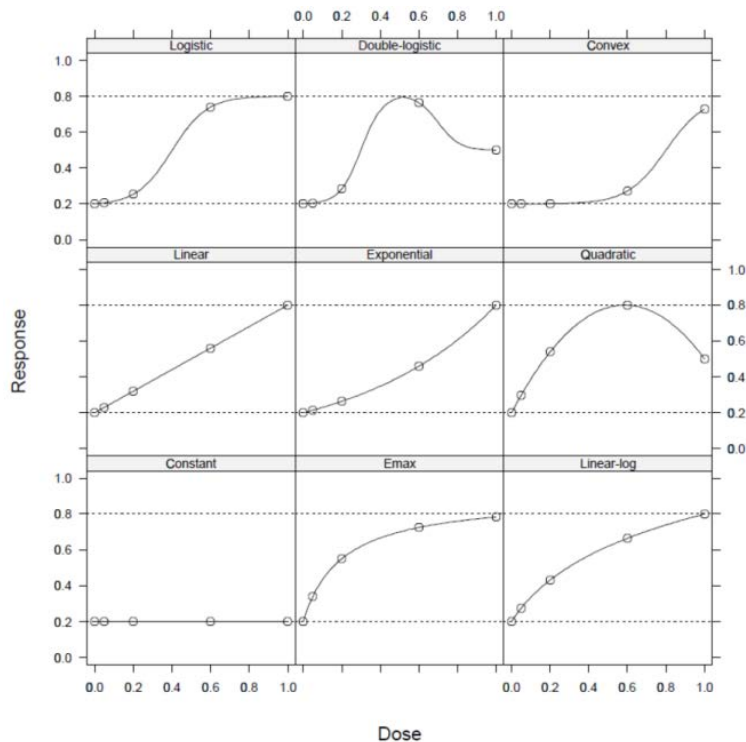
Source: Request for Qualification Opinion, Figure 3-7

Table 3:   **Dose-response Signal Detection Probabilities at Sample Size n=75 (Applicant's Simulation Study 1)**

Model Shape	MCP-Mod	LRT	Step Contrast
Constant	0.051*	0.052	0.055
E_{max}	0.868	0.857	0.841
Lin-log	0.891	0.874	0.858
Linear	0.880	0.865	0.850
Exponential	0.862	0.858	0.844
Quad	0.799	0.684	0.682
Logistic	0.960	0.946	0.944
Double Logist	0.805	0.721	0.726
Convex	0.728	0.761	0.753

Source: Adapted from Request for Qualification Opinion, Table 3-3.

Notes: Green, yellow and red cell color indicates the method with highest, second highest and lowest power.
* indicates the method with lowest type I error.

Reviewer's comments: Parts of simulation study 1 were independently replicated by the reviewers. Results are presented in section 6.

In addition to investigate the power of MCP-Mod in comparison to other methods, simulation study 1 investigated the probability of the MCP-Mod method to detect a "true" under simulation scenario pre-defined minimal effective dose. No comparison was performed to stepwise contrast method because it is not equipped to detect a minimal effective dose outside of the observed doses (no interpolation or extrapolation is possible). The LLR ratio test is designed to test the overall significance of a positive dose-response signal. Hence no comparison to select the minimal effective dose was made.

The minimal effective dose was defined as the smallest dose that reached the clinical relevant effect of 0.4 points over placebo. These simulations were conducted using the same model shapes as before but with a lower residual error (σ = 0.65, compared to σ =1.478). The minimal effective dose was estimated using 3 methods as shown in **Figure 2**. The three methods rely on the typical estimate, the upper, or the lower confidence interval with the estimated probability to reach the target effect size. Applicant reports using 80% confidence intervals for the purpose of finding the minimal effective dose.

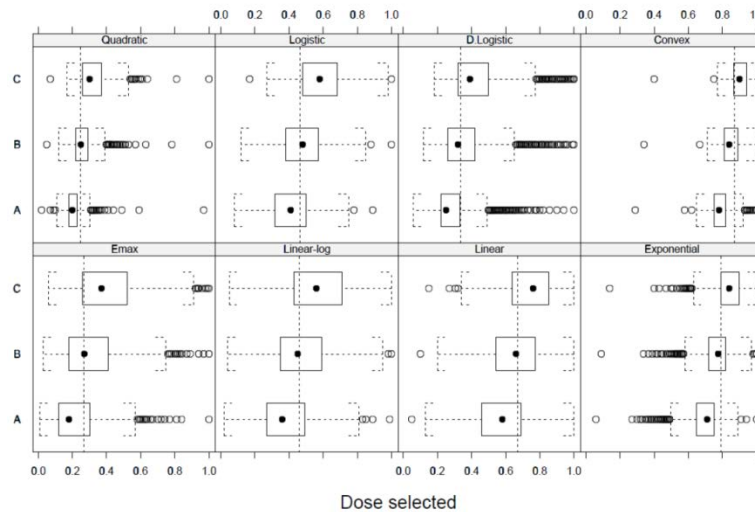**Figure 2:   Three Different Methods for Target Dose Estimation**
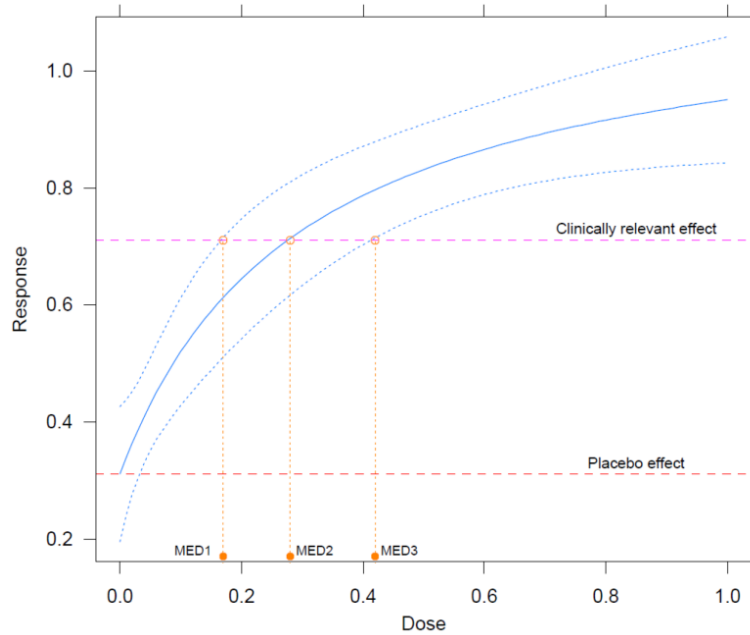
Result of applicant's simulations are shown in **Figure 3**. Notable finding are that the least unbiased estimate of the minimal effective dose comes from the use of the typical estimate of effect size (MED2).

Figure 3:   Boxplots of Selected Doses for Sample Size of n = 75

5.2.2 MCP-Mod versus ANOVA, BMA, and LOCFIT

Simulation Study 2

This simulation study was published in the white paper of the PhRMA working group on "Adaptive Dose-Ranging Studies" (Bornkamp et al. 2007). However, simulations presented here deal only with non-adaptive dosing methods. Applicant has submitted selected results from this simulation study. These are summarized below.

The ability to find the minimal effective dose was compared using an ANOVA based method, a Bayesian approach, a non-parametric approach, and finally MCP-Mod. These are described below.

> *"**ANOVA approach.** The particular ANOVA approach used in the simulation study consists of an initial one-sided Dunnett multiple comparison procedure to test each of the active doses against placebo. If at least one of the doses is statistically significant (under Dunnett's multiplicity adjustment), dose-response is established. The target dose is then estimated as the smallest statistically significant dose which has an average effect that is clinically relevant (according to a pre-specified value of clinical relevance), provided at least one dose meet both criteria. If a target dose can be estimated, the final step of the approach consists in estimating a dose-response model. Three candidate dose-response models (linear, quadratic, and logistic) are fitted to the data and the Akaike Information Criterion is used to select the best model, which is then used for predictions, etc."*
> Source: Request for Qualification Opinion. Page 45

Reviewer's comment: This is a hybrid approach that combined Dunnett's test with modeling approaches. Dunnett's test will lose power as more doses are tested.

> *"**Nonparametric dose-response modeling approach (LOCFIT).** This method relies on model- free testing techniques to assess a possible dose-response effect. Non-parametric regression techniques are used for target dose estimation as they can model virtually any smooth dose- response shape without the need to pre-specify a parametric dose-response model. The dose- response effect is assessed using a multiple contrast test. To cover a broad range of potential dose-response shapes, the method relies on five contrast tests capturing the concave, convex, sigmoid, linear, and umbrella model shapes (see Stewart and Ruberg, 2000 for more information on multiple contrast tests). For the dose estimation step we utilized local quadratic regression techniques (Hastie and Loader, 1993) using a Gaussian kernel and a global bandwidth. The bandwidth was selected by minimizing the generalized cross- validation score. The locfit package in R was used for the implementation of the procedure."*
> Source: Request for Qualification Opinion. Page 45

*"**Bayesian model-averaging approach (BMA).** Bayesian model averaging is a strategy intermediate between parametric modeling and nonparametric modeling that tries to avoid the dangers of under- or over-fitting. As a basis here a set of relatively simple dose-response models is used. Then, starting from prior model probabilities ("weights"), as well as prior distributions on the model-specific parameters, standard Bayesian inference leads to posterior updates of the unknown quantities (model weights and model parameters). The approach is well-suited for situations where the quantity of interest is model-independent, such as in dose- ranging studies where the objective is to find a dose fulfilling a certain pre-specified criterion. Bayesian model averaging generalizes model selection strategies and has the advantage of weighting the candidate models in an appropriate (data-dependent) way. A simple informal Bayesian model averaging approach based on a set of normal linear models allowing for analytic posterior updates was used. Here this choice was mainly dictated by the fact that MCMC-based posterior inferences for non-conjugate models would have made simulations computationally infeasible. The approach is informal in that it only uses posterior summaries as a basis for dose selections."*
Source: Request for Qualification Opinion. Page 45

Reviewer's comment: The original publication investigated the performance of other methods; these include Multiple Trend Test Approach (MTT), D-Optimal Response-Adaptive Approach (Dop), and Bayesian Adaptive Dose Allocation Approach (GADA). Applicant's decision to not include results from the methods that rely on adaptive trial design is understandable. It is however not clear why results from Multiple Trend Test approach were omitted since the MTT method is related to the MCP-Mod method and does not rely on adaptive trial elements.

Drug effect was simulated under different dose-response shapes (**Figure 4**), and various dosing scenarios.

- Five equally spaced doses: 0, 2, 4, 6, and 8.
- Seven unequally spaced doses: 0, 2, 3, 4, 5, 6, 8.
- All nine equally spaced doses: 0, 1, …, 8.

The maximum effect was set to -1.65 units for all model shapes. Residual error was assumed to be normally distributed with a mean of zero and variance of 4.5. Simulations were conducted with a total sample size of 150 or 250 patients. Total sample size is the sum of number of patients in each dose arm. Number of simulations per method varied between 5,000 and 10,000 samples.

*Reviewer's comments: These dosing scenarios differ significantly from the typical phase 2 trial in terms of number of doses and their spacing (**Table 1**). Traditionally dose-response trials do not have equal spacing between doses and usually do not include more than 3 to 4 active doses. The residual error variance of 4.5 or 2.12 standard deviations corresponds to a 0.78 signal to noise ratio. This ratio is larger than the signal to noise ratio in simulation study 1 part 1 (0.41) but smaller than the signal to noise ratio of simulation study 1 part 2 (0.92). The comparison*

of signal to noise ratios does not consider any other study design characteristics (number of doses, their spacing, and sample size).

The reviewer has noted that the applicant has chosen not to include simulations based on a sigmoid E_{max} model in the application. The reviewer will show these results as reported in the applicant's publication (Bornkamp et al. 2007) or in the comprehensive technical report[5].

Figure 4:   Dose-response Profiles Used in Simulation Study 2
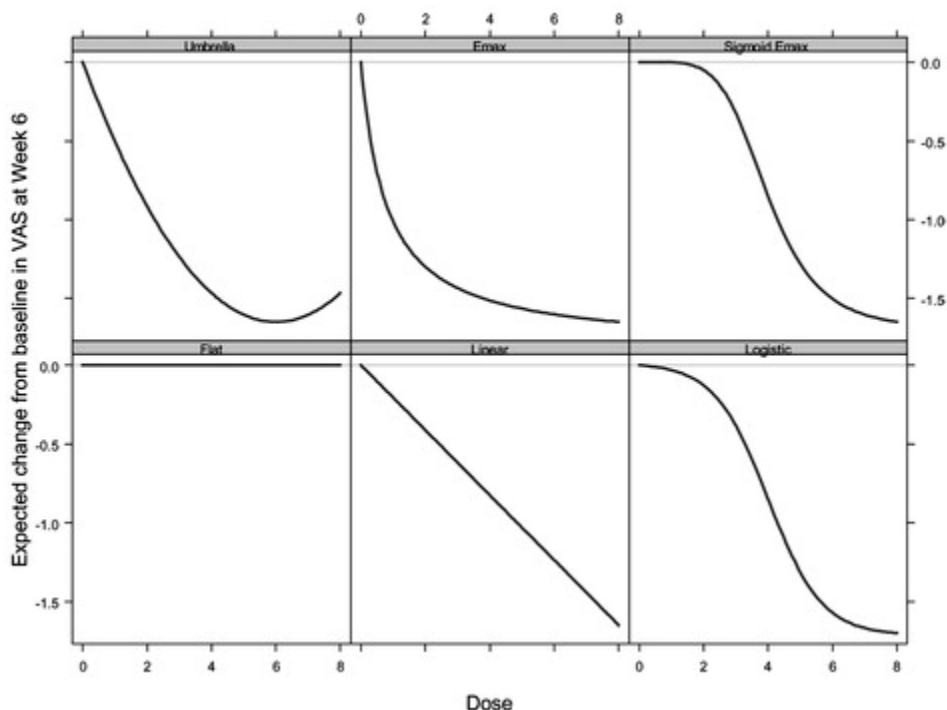
Source: Bornkamp et al. 2007.

The applicant investigated several performance metrics for the four methods. This reviewer will focus on a) power, b) type I error, c) and probability to identify the correct minimal effective dose.

5.2.2.1 Type I Error Rate Results

Probability of falsely declaring a dose-response signal is shown in **Figure 5.** All methods were able to control for type I error rate, although the MCP-Mod method was numerically worst in 4 out of 6 cases. The applicant attributes this to fluctuations attributed to Monte Carlo error.

Reviewer's comments: MCP-Mod type I error rates appear to be consistent with simulation study 1 part 1.

---

[5] http://www.biopharmnet.com/doc/doc12005-01.html

**Figure 5:   Type I Error Rate for Detecting Dose-response Under Flat
Dose-response Profile**
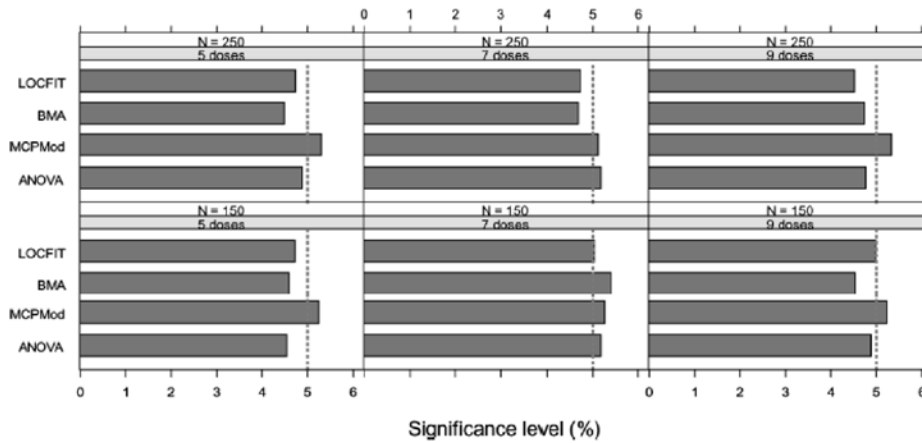
Source: Request for Qualification Opinion. Figure 3-12

5.2.2.2 Power to Declare Significant Dose-response

The probabilities of detecting true dose-response are shown in **Figure 6**. As expected the ANOVA followed by Dunnett's test loses power when the number of doses increase due to adjustment for multiplicity. The other three methods appear to consistently outperform ANOVA.

The MCP-Mod approach and the other two approaches were also shown to lose power when number of doses increase from 5 to 9. The applicant acknowledges this finding in their 2007 publication but not in the current submission.

> *"It is worth noting that increasing the number of doses, while keeping the total sample size fixed, does not lead to an increase in power to detect DR, even for model-based approaches. In fact, for some model-method combinations, Pr(DR)decreased with the number of doses."*
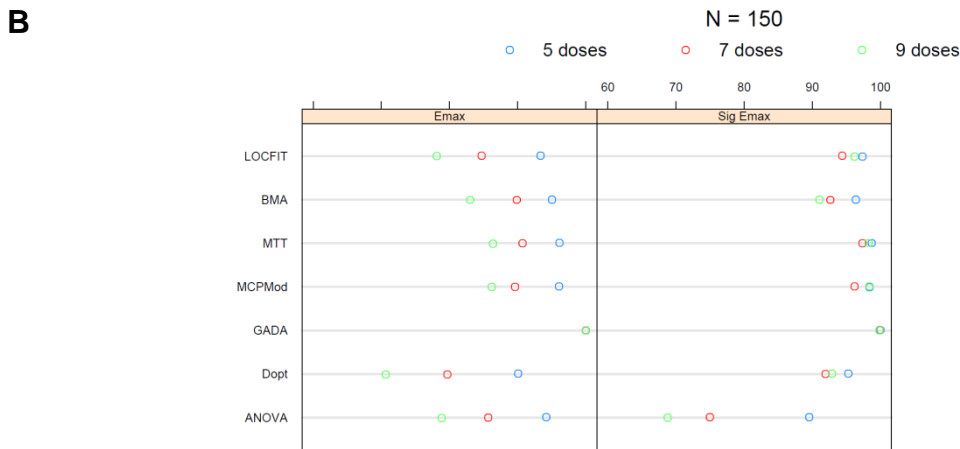>
> *Bornkamp et al. 2007*

Results from simulations using the sigmoid E_{max} model were omitted in this submission. They have been retrieved by the reviewer and presented in **Figure 6 B.** It is noteworthy that MCP-Mod method performs better under the sigmoid E_{max} model.

Figure 6:   Power to Detect Dose-response Under Active Dose-response Profiles.
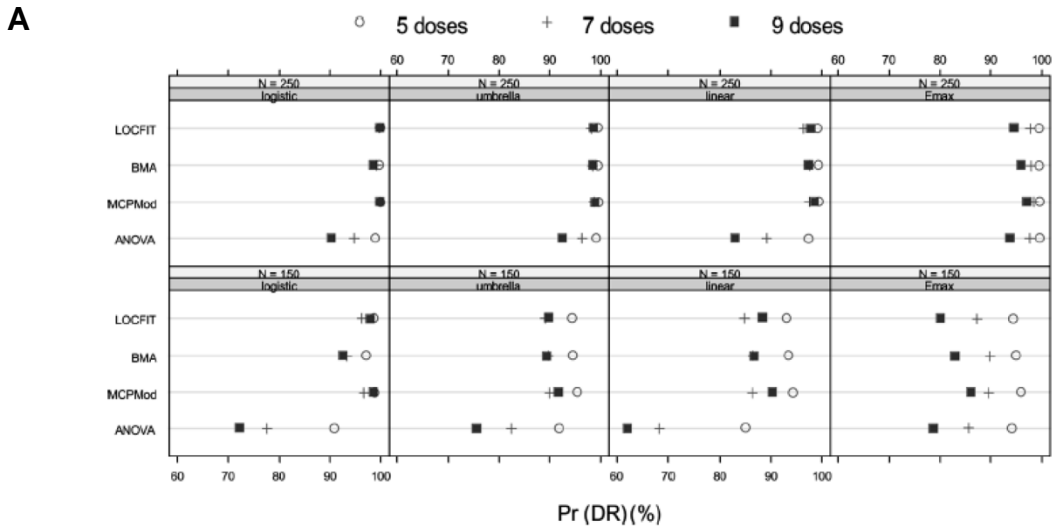
A

B

5.2.2.3 Probability to Identify Minimal Clinical Effective Dose

The smallest dose that results in efficacy above the presecified clinical relevant level (−1.3 units) is selected as the minimal effective dose. The selected dose is rounded up to the nearest integral of 1, 2…8. A dose not studied in the trial could be selected as the minimal effective dose.

Probability to identify or recover the predefined minimal effective dose for each of the 4 methods under the different dose-response shapes is shown in **Figure 7 A.** Probability for the sigmoidal E$_{max}$ model is shown in **Figure 7 B.**

The MCP-Mod method appears to have similar performance as the other methods. Results are dependent on sample size and the underlying model shape. The traditional ANOVA approach is used here with additional modeling steps. Models used in the ANOVA approach were limited to linear, quadratic, and logistic. Under those model shapes and at the larger sample size, the ANOVA method coupled with modeling appears to perform as good as or better than the other methods. It is reasonable to think that if the ANOVA method were to include other model shapes, performance would improve. However, the underlying drawback of ANOVA method (loss of power as more doses are tested) is not resolved by adding more model shapes.

As previously noted, performance appears to decrease as more doses are studied. This finding is again observed for this performance metrics.

Figure 7: Probabilities of Identifying Clinical Relevant Dose Under Active Dose-response

A
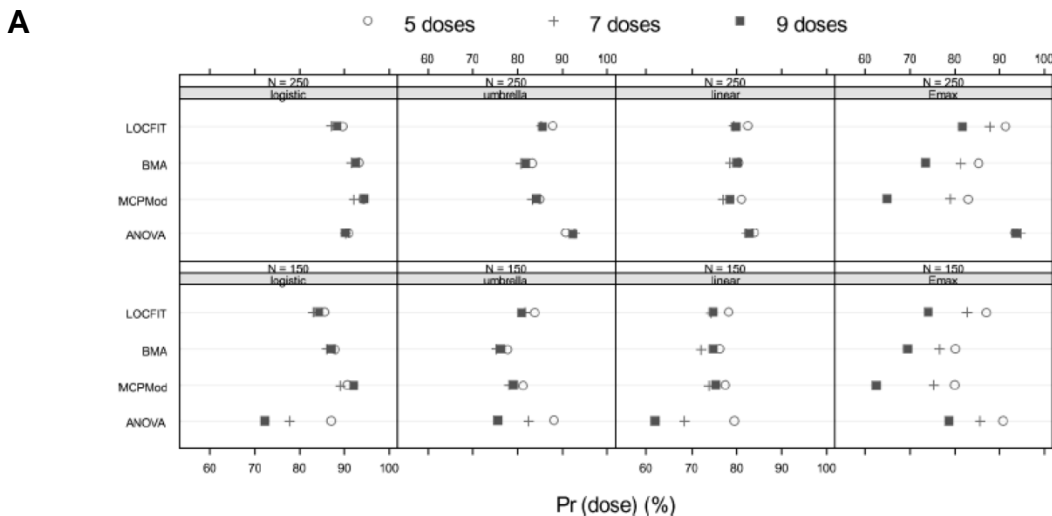
Source: Request for Qualification Opinion. Figure 3-15

B
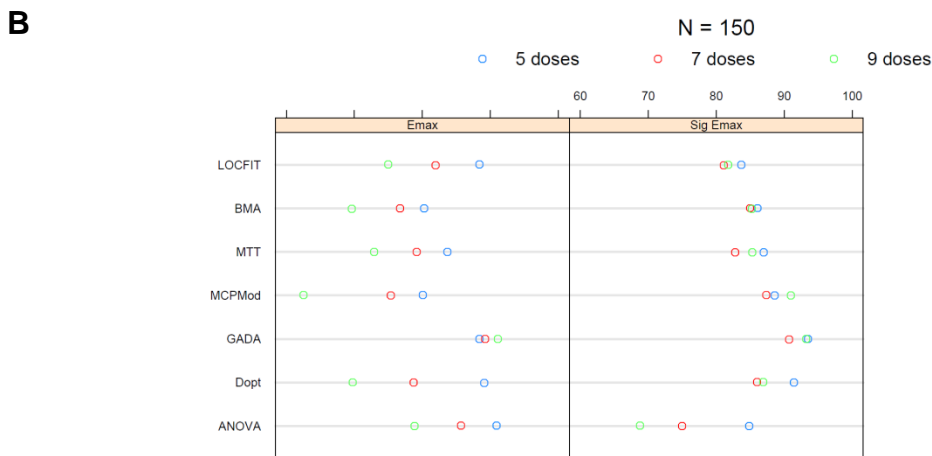
Source: Technical report. http://www.biopharmnet.com/doc/doc12005-01.html
Note: The pane showing E_{max} results is identical to the panel in figure 7 A. Only non-adaptive methods are considered in this review. Only non-adaptive methods are considered in this review.

5.2.3 Reviewer's Conclusions

The applicant submitted an overview of two simulation studies where MCP-Mod was assessed based on metrics relevant to the analysis of dose ranging trials. These metrics included type I error rate, power to detect a significant dose-response shape, and the power to find the minimal effective dose. Comparisons were made to other data analysis methods; these included a combination of ANOVA with Dunnett's test coupled with a limited modeling approach; a Bayesian modeling approach, and a non-parametric modeling approach. MCP-Mod was also compared to the log likelihood ratio test (linear regression) but only in terms of power and type I error rate.

The simulation studies were conducted in a manner that explores different model shapes and sample sizes. However based on literature reports on phase 2 study design, see section 4, and this reviewer's experience, there is a disconnect between what is the typical phase 2 study and the simulations reported here. Particularly in the sense that a typical phase 2 study includes 3 to 4 dose while the number of doses in applicant's simulations ranged between 5 and 9. As previously stated (see section 4.1), it is reported in literature that dose ranging data typically has an E_{max} type of shape. The E_{max} shape was indeed investigated in these simulations; however other shapes (even some very unlikely shapes) are given equal weight in these simulations. This reviewer is therefore making his conclusions with special attention to simulations that are closest to the typical dose ranging trial.

With respect to power and type I error rate, the ANOVA based method suffered (as expected) from loss of power with increasing number of doses. So did the other methods (including MCP-Mod) but to a less extend than ANOVA. MCP-Mod method was not shown to be superior nor was it inferior to LLR test, Bayesian modeling or non-parametric modeling.

The LLR test and MCP-Mod had very similar performance. Both methods where shown to have their weakness and strengths. The MCP-Mod method seems to be sensitive to the initial 'guesses' of the model shapes that are included in the analysis. This was evident in simulation study 1 part 1 when power of the MCP-Mod approach was significantly lower when the data was simulated with a model not included in the MCP-Mod library. A somewhat counter-intuitive finding was that MCP-Mod and other modeling methods performed worse when number of dose arms increased but the total sample size remained the same. A possible explanation to this finding is that the design with smaller number of doses included more subjects at highly optimal or influential doses compared to the other designs. Reviewer's own simulation study lends support to this explanation, see section 6. It should be noted that both the model shape and model parameters need to be known *a priori* in order to design a dose-ranging study based on the D-optimal design. This is almost never the case in clinical drug development. It is therefore good practice to explore a wide range of doses so that the shape of the dose response curve can be characterized.

Power to find the minimal effective dose differed across model shapes and sample sizes. No method was consistently better than the other.

6   REVIEWER'S ANALYSIS

Introduction

The purpose of this simulation study is to compare the MCP-Mod procedure to the log-likelihood ratio test. Log-likelihood ratio test is widely used to assess the global dose (exposure) response relationship and is currently recommended in published best practice principals of exposure-response analysis (Overgaard et al. 2015).

Reviewer's analysis was conducted in collaboration with Dr. Lei Nie (Office of Biostatistics).

Objectives

The objective of this analysis is to confirm the applicant's finding with an independent analysis. The analysis will focus on the following:

Objective 1:   Assess the power of the MCP and the LLR test to detect a true dose-response relationship under different simulated true dose-response relationships.

Objective 2:   Assess type I error rate of the MCP and LLR when no true dose-response effect is present.

Objective 3:   Assess the influence of increasing the number of doses when the total samples size is the same on power to detect true dose response relationship.

Objective 4:   Asses the influence of using D-optimal design on power to detect true dose-response relationship.

Methods

The maximal observed effect adjusted for placebo was 455 under all model shapes. The dose resulting in the highest effect was the highest dose simulated for all models except for the Umbrella model where the nominal dose of 50 resulted in highest effect over placebo, **Figure 8**. All models were simulated with a residual error with the mean of 0 and the standard deviation of 400. Simulations were repeated 10,100 times. A power of 90 % (α=0.05) is achieved at a sample size of 18 for the one sided pairwise test at the maximum effect over placebo (455 units).

The following doses were tested in the simulation study

Objectives 1 and 2. Effect at placebo (dose: 0), and nominal doses of 25, 50, and 100 were simulated for all models. This design is considered to be the 'standard design' of phase 2 dose ranging trials.

Objective 3. Nine (9) doses including placebo were simulated. These were: 0, 2.5, 5, 10, 20, 40, 60, 80, and 100.

Objective 4. Tree doses (0, 10, 100) were selected based on the D-optimal design under the E_{max} model. Optimal doses were restricted to be between 0 and 100 and to increase in increments of 5 units.

Figure 8: **Model Shapes Used for Simulation. The Observed Doses are Shown as Points. Lines Connecting the Points Show the Underlying Dose-Response Relationship**
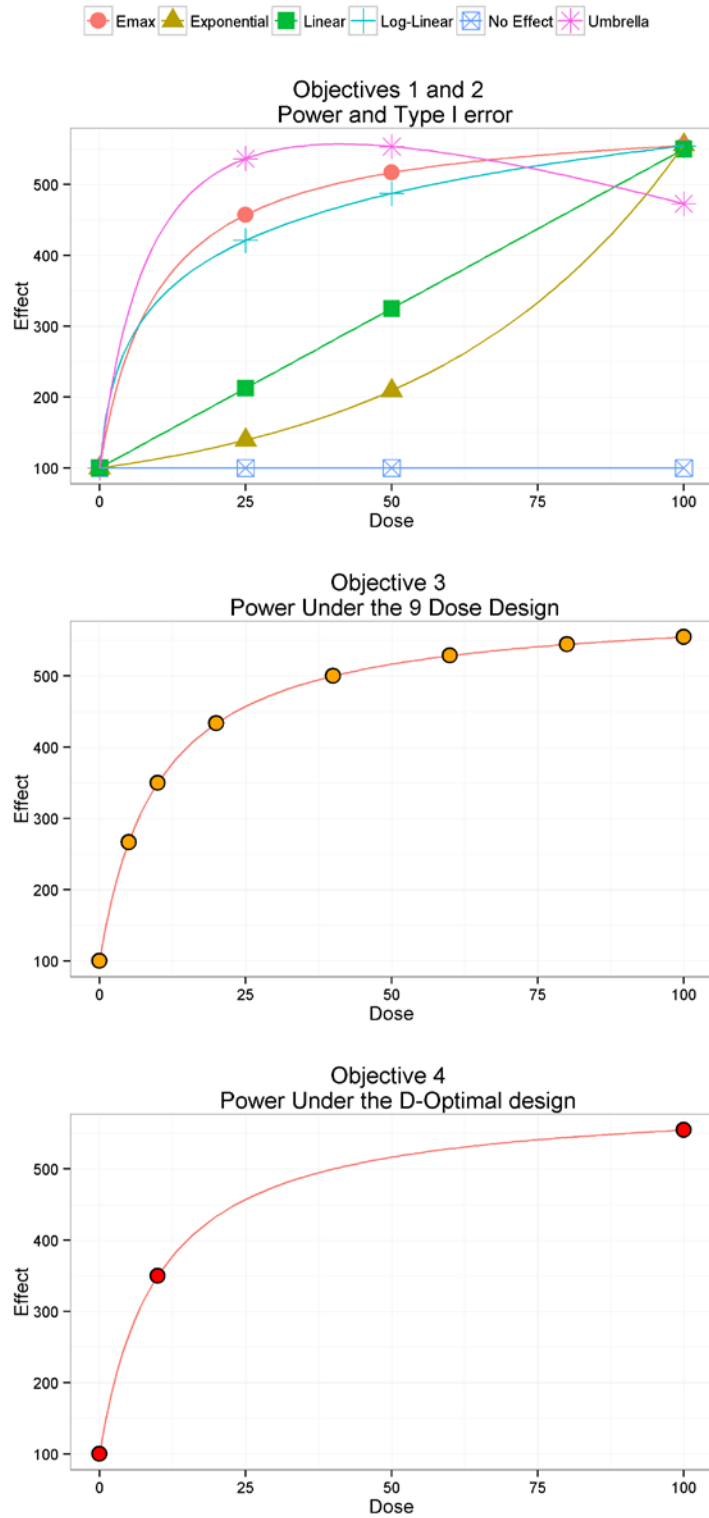
Note: Figures generated by the reviewer with R code
"Simulated_models.R"

Methods: Software

R (version 3.2.1) in conjunction with R Studio (version 0.99.486) was used for simulation, data analysis and visualization. DoseFinding (version 0.9-13) package in R was used to find the D-optimal design and to apply the MCP-Mod method.

Methods: Models

Model equations and model parameters used in these simulations are shown in **Table 4.**

Table 4. **Model Equations and Parameter Values Used in Reviewer's Simulations**

Model	Equation	Parameter definition and numerical value used in simulations
E$_{max}$	$E = E_0 + \dfrac{E_{max} \times Dose}{ED_{50} + Dose} + \sigma$	$E_{max}: Maximal\ effect\ (500)$ $ED_{50}: Dose\ requared\ to\ reach\ E_{max}/2\ (10)$ $E_0: Placebo\ response\ for\ all\ models\ (100)$ $\sigma: residual\ error, mean\ of\ 0\ and\ sd\ of\ 400$ $for\ all\ models$
Linear	$E = E_0 + \delta * Dose + \sigma$	$\delta: Slope\ parameter\ (4.5)$
Log-Linear	$E = E_0 + \delta * \ln(Dose + 1) + \sigma$	$\delta: Slope\ parameter\ (98.5)$
Umbrella	$E = E_0 + \dfrac{E_{max} \times Dose}{ED_{50} + Dose} - q \times Dose + \sigma$	$E_{max}: Maximal\ effect\ if\ q = 0\ (705)$ $ED_{50}: Dose\ requared\ to\ reach\ E_{max}/2\ (10)$ $q: negative\ slope\ (2.68)$
Exponential	$E = E_0 + E_1 * \exp\left(\dfrac{Dose}{\delta} - 1\right) + \sigma$	$\delta: convexity\ parameter\ (43.21)$ $E_1: Slope\ paramete\ (50)$

Results

Objectives 1 and 2: Power and Type I Error Rate

Results are visualized in **Figure 9.** The type I error rate was similar and acceptable for both MCP-Mod and LLR. A small numerical difference was observed in favor of MCP-Mod.

MCP-Mod method and the LLR test have similar power to detect a significant dose-response relationship for most model shapes. LLR test was numerically superior for the linear and the exponential model while MCP-Mod was superior under the E$_{max}$ and the log-linear model shape. As expected power of the LLR test is significantly lower under the umbrella model shape because of the monotonicity assumption of the LLR test.

Both the MCP-Mod procedure and the LLR[6] test had >90% power, which is higher than the power for the pairwise comparison at the dose with the highest effect.

These results are similar to what the applicant reports in their simulation study. Both MCP-Mod and LLR test have assumptions that when violated result in lower power to detect dose-response relationship. In case of the LLR test, that assumption is monotonicity of the dose-response shape. The assumption of MCP-Mod is that the observed model shape was included as one of the prespecified models. Applicant's simulations show that power decreases significantly when that assumption is violated, see section 5.2.1 and **Table 3.**

Objectives 3 and 4: Power Under the 9 Dose Design and Under the D-optimal design

As shown in **Figure 9,** power is ~10% higher under the D-optimal design which only includes 3 doses compared to the 9 dose design (total sample size is the same for both designs). MCP-Mod power was very similar under the D-optimal design and the standard 4 dose design. This is likely due to the similar choice of doses in the two designs, see **Figure 8**. If the standard design included doses only on the flat part of the curve it is reasonable to expect that power of MCP-Mod would be significantly lower compared to the D-optimal design. LLR test performed numerically better than the MCP-Mod procedure under the 9 dose design. The two methods were equally good under the D-optimal design.

---

[6] This is not true for the LLR test under the Umbrella model where power is estimated to 68.44%

Figure 9: Type I Error and Power Under Different Study Designs and Model Shapes. (Reviewer's Simulation Study)
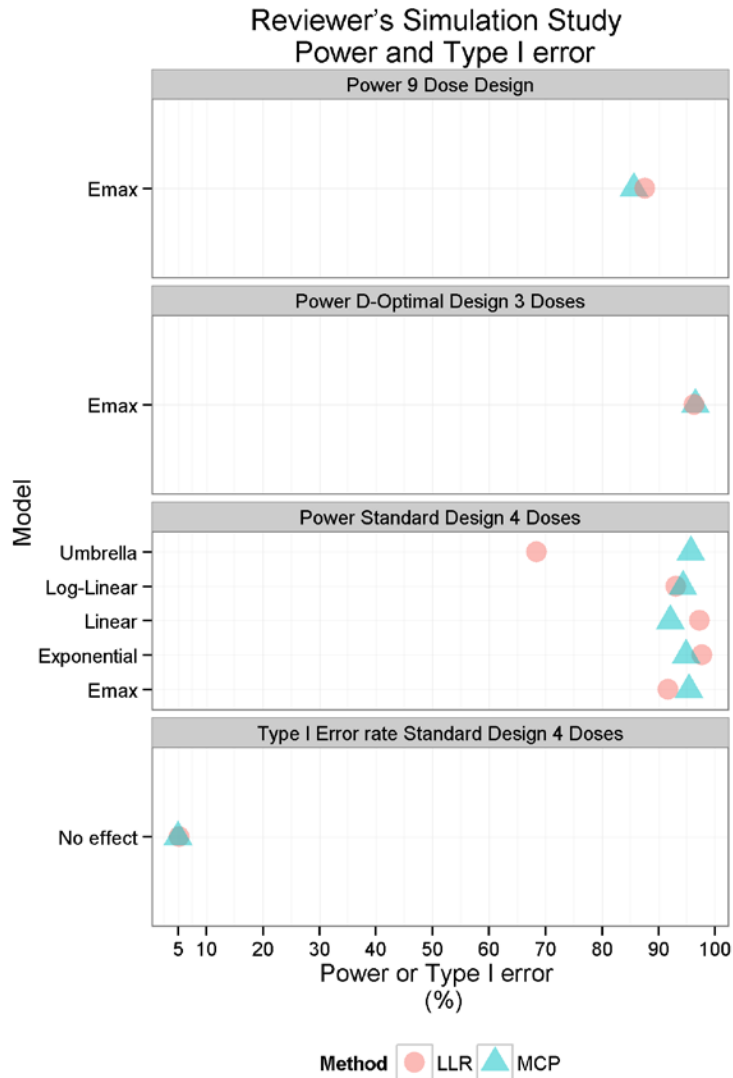
Note: Total sample size was 72 for all scenarios. All simulations where repeated 10,000 times. Maximal observed effect over placebo was same for all scenarios (450 units). Residual error was also the same for all scenarios (SD=400).

Listing of Analyses Codes and Output Files

File Name	Description	Location in \\cdsnas\pharmacometrics\
D_optimal_emax.R	R code used to calculate optimal design for the E_{max} model in the reviewer's simulation study	…/MCP_MOD_DR
Simulated_models.R	R code used to generate **Figure 8**	
Power.T.test.R	R code used to calculate samples power in reviewer's simulation study	
Sim.Eff.R Sim.Eff_analysis	R code used to estimate power under the E_{max} model in reviewer's simulation study	
Sim.Eff_expo.R Sim.Eff_expo_analysis.R	R code used to estimate power under the Exponential model in reviewer's simulation study	
Sim.Eff_linear.R Sim.Eff_Linear_analysis.R	R code used to estimate power under the Linear model in reviewer's simulation study	
Sim._LOG_lin_Eff.R Sim.Eff_Log_Linear_analysis.R	R code used to estimate power under the Log-Linear model in reviewer's simulation study	
Sim.Eff_umbrella.R Sim.Eff_Unbrella_analysis.R	R code used to estimate power under the Umbrella model in reviewer's simulation study	
Sim.No.Eff.R Sim.No.Eff_analysis.R	R code used to estimate type I error rate in reviewer's simulation study	
Results.plot.R	R code used to generate **Figure 9**	
Sim3doses.Eff.R Sim3doses.ANNALYSIS Sim9doses.Eff.R Sim9doses.ANNALYSIS	R code used to estimate power under the E_{max} model based on D-optimal design or 9 dose design.	
D_optimal_emax.R	R code to estimate the D-optimal design.	

7 REFERENCES

Aoki, Yasunory, Bengt Hamrén, Daniel Röshammar, and Andrew C Hooker. 2014. "Averaged Model Based Decision Making for Dose Selection Studies." http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2015/01/WC500179793.pdf. Retrieved 12/01/2015.

Bartholomew, D. J. "Ordered tests in the analysis of variance." Biometrika (1961): 325-332.

Bauer, P., and P. Hackl. "The application of Hunter's inequality in simultaneous testing." Biometrical journal 27.1 (1985): 25-38.

Bornkamp, Björn, et al. "Innovative approaches for designing and analyzing adaptive dose-ranging trials." Journal of biopharmaceutical statistics 17.6 (2007): 965-995.

Branson, Michael, José Pinheiro, and Frank Bretz. "Searching for an adequate dose: Combining multiple comparisons and modeling techniques in dose-response studies." (2003). Novartis Biometrics Technical Report No. 2003-08-20. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=1A7992413E16CB9CA4295A9FF07B6EB5?doi=10.1.1.201.1944&rep=rep1&type=pdf. Retrieved 12/1/2015

Cross, James, et al. "Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999." Pharmacoepidemiology and drug safety 11.6 (2002): 439-446.

Dunnett, Charles W. "A multiple comparison procedure for comparing several treatments with a control." Journal of the American Statistical Association 50.272 (1955): 1096-1121.

EMA."Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty" (2014) http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2014/02/WC500161027.pdf. Retrieved 12/2/2015

FDA. 2009. "Guidance for Industry: End-of-Phase 2A Meetings."

Filloon, Thomas G. "Estimating the minimum therapeutically effective dose of a compound via regression modelling and percentile estimation." Statistics in medicine 14.9 (1995): 925-932.

ICH-4. 1994. "Dose-response Information to Support Drug Registration."

Lalonde, R. L., et al. "Model-based drug development." Clinical Pharmacology & Therapeutics 82.1 (2007): 21-32.

Overgaard, R. V., S. H. Ingwersen, and C. W. Tornøe. "Establishing Good Practices for Exposure–Response Analysis of Clinical Endpoints in Drug Development." CPT: pharmacometrics & systems pharmacology 4.10 (2015): 565-575.

Sacks, Leonard V., et al. "Scientific and regulatory reasons for delay and denial of FDA approval of initial applications for new drugs, 2000-2012." JAMA 311.4 (2014): 378-384.

Sheiner, Lewis B. "Learning versus confirming in clinical drug development*." Clinical Pharmacology & Therapeutics 61.3 (1997): 275-291.

Thomas (a), Neal, Kevin Sweeney, and Veena Somayaji.. Meta-analyses of clinical dose-response. Presented at EMA/EFPIA, 2014, http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2015/01/WC500179795.pdf. Retrieved 12/1/2015

Thomas (b), Neal, Kevin Sweeney, and Veena Somayaji. "Meta-Analysis of Clinical Dose–Response in a Large Drug Development Portfolio." Statistics in Biopharmaceutical Research 6.4 (2014): 302-317.

Williams, D. A. "A test for differences between treatment means when several dose levels are compared with a zero dose control." Biometrics (1971): 103-117.

Williams, D. A. "The comparison of several dose levels with a zero dose control." Biometrics (1972): 519-531.

Yang, Jun, et al. "The Combination of Exposure-Response and Case-Control Analyses in Regulatory Decision Making." The Journal of Clinical Pharmacology 53.2 (2013): 160-166.