Design and analysis of pre-clinical experiments using a method combining multiple comparisons and modeling techniques for dose-response studies

Avijit Singh

Abstract. Identifying and estimating the dose-response relationship between a compound and a pharmacological endpoint of interest is one of the most important and difficult goals in the preclinical stage of pharmaceutical drug development. We conduct pharmacodynamic studies to investigate the dose-response profile and then different studies to find doses that lead to desired efficacy or acceptable safety in the endpoint(s). The aim of this thesis is to provide an overview of existing techniques and design of experiments which are appropriate for addressing the goals of these studies simultaneously. We have used a method combining multiple comparisons and modeling techniques (MCPMod) in designing the experiments and found that we can reduce the required total sample size by using an optimal design. We have analysed the simulated data using MCPMod and observed that this method can be used to identify the dose-response relationship and estimate dose at a required effect. We have compared the two approaches of estimating dose and discovered that using a weighted average of all fitted models gives a similar result as compared with using the best fitted model. Finally we have investigated the possibility of identifying the presence of toxicity in response of a few or many samples at higher doses and found that we can detect toxicity if there are many samples with toxic response at a higher dose. This combined strategy is both financially and ethically rewarding as it reduces the time and cost of study and also reduces the number of animals used in pre-clinical trials.

Key words: pharmacodynamic modeling; dose–response relationship; significance level; multiple comparison procedure; optimal design; MCPMod; dose finding study

UNIVERSITY OF GOTHENBURG

AstraZeneca

Date: 2021/03/25
Thesis for the degree of Master of Science, written in the field of mathematical statistics as part of the two year *Master's Programme in Mathematical Sciences* at the University of Gothenburg in Sweden. Please direct correspondence to gussinav@student.gu.se or avijit.iitk@gmail.com

Supervisor: Professor Umberto Picchini, Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg
Supervisor: Peter Konings (Associate Director Statistics, AstraZeneca AB)
Co-advisor: Pär Nordell (Preclinical and Translational PK & PKPD,AstraZeneca AB)
Examiner: Professor Marina Axelson-Fisk, Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg

Acknowledgements

I would like to express my gratitude to Prof. Umberto Picchini and the company AstraZeneca AB(AZ) for giving me the opportunity to write this thesis in cooperation with the Department of Mathematical Sciences at the University of Gothenburg.

Many thanks to Prof. Umberto Picchini for taking over the supervision of my thesis. I appreciate your guidance and engagement throughout the whole time.

My supervisor from AZ, Peter Konings, deserves special thanks for providing me the opportunity to work on this important and interesting topic and for giving me the chance to write my thesis at AZ. Your excellent supervision and all your suggested improvements helped me a lot. Furthermore, my gratitude to Pär Nordell for the discussions and suggestions. In addition, I want to thank the whole Quantitative Biology Statistics group at AZ for the warm welcome and the offered support whenever needed and special thanks to Tom Marlow for reviewing the report and providing valuable inputs.

Finally, I would like to thank my family and my friends for the encouragement and support during the time of studying, particularly during the work on this thesis.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Drug discovery and development is a highly challenging process and identification of the appropriate dosage is a one of the most important steps in the development process of a new drug. During pre-clinical testing, 75% of safety closures were compound-related (that is, they were due to 'off-target' or other properties of the compound other than its action at the primary pharmacological target) as opposed to being due to the primary pharmacology of the target.[1] If the target dose is selected too high it can result in an unacceptable toxicity profile. Alternatively, if the target dose is selected too low it increases the chance that the compound provides insufficient evidence of effectiveness.[2]

Searching for an adequate dose has historically been addressed using two approaches: multiple comparisons and modeling. The former regards the dose as a qualitative factor and generally makes few, if any, assumptions about the underlying dose-response relationship. The latter assumes a functional relationship between the response and dose, taken to be a quantitative factor, according to a pre-specified model.[3] The first approach of multiple comparisons typically has relatively few dose-levels, with many observations made at each level in order to yield statistical confidence in the results. The second approach of multiple comparisons typically has a larger number of dose-levels, spread out in such a way that we gain insight into the functional relationship between dose and response, whilst sacrificing confidence by having fewer observations at each dose-level. Regularly, however, an optimal design for either of these endpoints might be sub-optimal for the other (e.g. number of drug dose levels vs number of animals in each group), which can necessitate follow-up studies with additional use of animals.

A unified strategy is possible to analyse data from dose-response studies which combines multiple comparison and modeling techniques. This assumes the existence of several candidate parametric models and uses multiple comparison techniques to choose the one most likely to represent the true underlying dose-response curve, while preserving the family-wise error rate. The selected model is then used to provide inference on adequate doses.[3]

The goal of this thesis to understand this unified strategy which can address efficacy and dose-response endpoints simultaneously and then apply the above strategy on multiple designed experiments. A combined strategy is rewarding both financially and ethically as it reduces the time and cost of study and also reduces the number of used animals in pre-clinical trials.

A pioneering analysis using benchmark industry attrition data in 2004 highlighted the most common causes of R&D attrition as being a lack of efficacy and/or safety and suggested that more attention be spent on reducing toxicity risks, improving pre-clinical models and demonstrating adequate proof of mechanism and proof of concept in the clinic.[4] This study also investigate that using the above mentioned combined strategy we can get more insights about the efficacy and dose-response end-points during pre-clinical experiments.

1.2 Pharmacological Background

All healthcare companies are making substantial investments in their capabilities for target selection and validation, lead generation, pharmacokinetic/pharmacodynamic (PK/PD) modeling, patient stratification and biomarkers. Dose selection and dose regimen design are essential for converting drugs from chemicals to therapeutically useful agents.[5]

PK models describe the relationship between drug concentration(s) and time. A primary goal of most population pharmacokinetic modeling evaluations is finding pharmacokinetic parameters and sources of variability in a population. Other goals include relating observed concentrations to administered doses through identification of predictive covariates in a target population.[6]

Models describing continuous PD metrics often represent the concentration–effect relationship as a continuous function. PD modeling has been shown to be important during regulatory review of new therapeutics.[7]

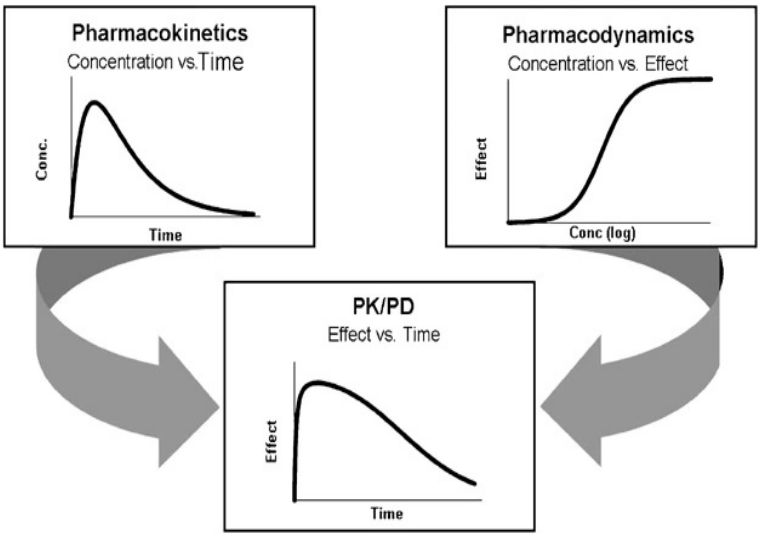

Figure 1: PK/PD modeling as combination of PK and PD [11]

Baseline (predrug) and control (vehicle dose) data are as important for PD modeling as the drug treated data and should be given equal attention. It is only with a faithful model of baseline and control data that it is possible to distinguish drug effect from control effect and from noise in the data.[8] This thesis focuses on dose–response studies, in which a drug is administered at various dose-levels and the corresponding response is measured. The medicinally active substance is delivered in an inert substance (excipient) referred to as vehicle.This vehicle treatment is a control expected to have no effect, and as such, it can be thought of as a type of control. Since the vehicle treatment contains no trace of the active substance, it is appropriate to associate it with a dose-level of zero.

Most PD models are semi-physiologic and are parameterized to reflect what is known about the drug's mechanism of action. Despite this, identification of appropriate structural models for PD evaluations is not straightforward. Comparing models with a linear drug effect vs. a nonlinear drug effect are not nested (such that fixing one new parameter to 0 simplifies the test model to the reference linear model), thus the likelihood ratio test is generally not an appropriate metric to compare different models. Instead, the use of the Akaike Information Criteria(AIC) or similar comparisons should be used.[6] It is to be noted here that AIC is not suitable when we want to predict the observations using fitted model, in our case we will not predict the data using the fitted models so this can be used. Preference is generally given to models that converge and provide robust parameter estimates with respect to initial selected candidate models and their parameters.[8] In our designed experiments, we have compared the results of estimated dose by using the best fitted model and by taking weighted average of all fitted models.

All drugs exhibit between-subject variability (BSV) in exposure and response, and many studies performed during drug development are aimed at identifying and quantifying this variability.[9] In this thesis we have assumed the between subject variability is normally distributed with given mean and response for a particular dose.

Integrated PK/PD modeling allows a quantitative description of the dose–exposure–response relationship by linking the concentration–time course resulting from a dosing regimen as assessed by PK to the intensity of observed response in relation to drug plasma concentrations as quantified by PD. The resulting PK/PD models allow a description of the complete time course of the intensity of desired and/or undesired effects resulting from a certain dosing regimen.[10]
Figure 1 illustrates the PK/PD modeling as combination of PK and PD.
In this thesis our focus is on PD models, which represent concentration(dose)–effect(response) relationship as a continuous function.

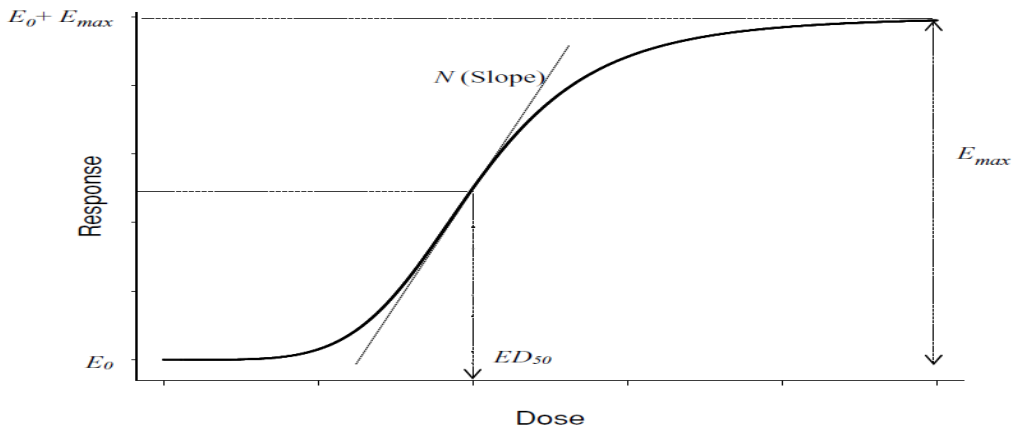

Figure 2: Sigmoid Emax model dose-response curve [12]

2 Modeling and Data Simulation

2.1 Mathematical Models

Throughout this thesis we will consider the functional relationship between a dose variable d and a response variable Y. Denoting the response function by f we will write Y = $f(d, \theta) + \epsilon$ where θ is a vector containing all parameters of the given response function and ϵ is random error.

The Emax model is a nonlinear model frequently used in dose–response analyses.[12] The model can be shown as:

$$Y = f(d, \theta) + \epsilon = E_0 + \frac{d \cdot E_{max}}{d + ED_{50}} + \epsilon$$

Where Y is the response, d is the dose, and parameters for this model are $\theta = (E_0, E_{max}, ED_{50})$ and E_0 is the value of response at control or zero dose, E_{max} is the maximum effect attributed to the drug, and ED_{50} is the dose, which produces half of E_{max}.
Notice that E_{max} is the change in response relative to the baseline E_0, meaning that the absolute maximal response is y = E_0 + E_{max}

If response is decreasing with the increase in dose, the Emax model can be written as follows:

$$Y = f(d, \theta) + \epsilon = E_0 - \frac{d \cdot E_{max}}{d + ED_{50}} + \epsilon$$

Where d, Y , and θ are the same as in the increasing Emax model.

The more general models of the Emax model is called the Sigmoid Emax model, which includes a slope factor (Hill factor), represented here as N, measures sensitivity of the response to the dose range of the drug, determining the steepness of the dose–response curve. The model is written as follows:

$$Y = f(d, \theta) + \epsilon = E_0 + \frac{d^N \cdot E_{max}}{d^N + ED_{50}^N} + \epsilon$$

Where all other variables are same as above except that θ has one additional parameter which is N, so $\theta = (E_0, E_{max}, ED_{50}, N)$

Figure 2 illustrates the Sigmoid Emax model dose-response curve where the response is increasing with the dose. We can see that for the particular case of Sigmoid Emax, when N =1, the model is referred as the Emax model. To cover the different shapes of dose-response relationships in the thesis, we will also use a logistic model and a beta model for the dose-response relationship. The

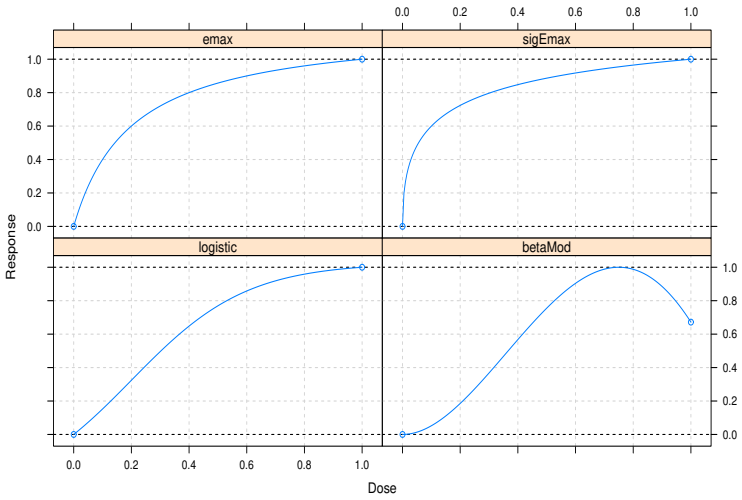

Figure 3: Dose-Response curve for all described models, For all models dose range is [0,1] , and parameters for standardized models are: for Emax: ED_{50} =.2 , For SigEmax: ED_{50} =.2 and N = 0.5, For logistic: ED_{50} = 0.2 and δ = 0.2, For Beta model: δ_1 = 2, δ_2 = 1.2, and scal = 1.2

logistic model is intended to capture general monotone, sigmoid dose-response relationships and the beta model is intended to capture non-monotone dose-response relationships. Using the beta model we can capture the case when the peak of the response occurs before the highest dose level and then decreases in the dose-response relationship. This could be an indication that there is a toxic effect at higher doses and we will use a beta model to detect the toxicity. Both the logistic and the beta model are considered as given in the R[13] package *DoseFinding* [14].

The logistic model is as follows: $Y = f(d,\theta) + \epsilon = E_0 + \frac{E_{max}}{1+\exp\frac{ED_{50}-d}{\delta}} + \epsilon$

Where Y is the response, d is the dose, and parameters for this model are $\theta = (E_0, E_{max}, ED_{50}, \delta)$, E_0 and E_{max} are the same as defined above, δ is the parameter determining the steepness of the curve. .

The Beta Model is as follows: $Y = f(d,\theta) + \epsilon = E_0 + E_{max} \cdot B(\delta_1, \delta_2) \cdot (\frac{d}{scal})^{\delta_1} \cdot (1 - \frac{d}{scal})^{\delta_2} + \epsilon$

Where Y is response, d is dose, and parameters for this model are $\theta = (E_0, E_{max}, \delta_1, \delta_2)$, E_0 and E_{max} are the same as defined above, δ_1 and δ_2 are shape parameters. $B(\delta_1, \delta_2) = \frac{(\delta_1+\delta_2)^{\delta_1+\delta_2}}{(\delta_1)^{\delta_1}\cdot(\delta_2)^{\delta_2}}$ and the kernel of the beta model function consists of the kernel of the density function of a beta distribution on the interval [0,scal]. The parameter scal is not estimated but set to a value larger than the maximum dose via the argument scal.

It is to be noted that all dose-response models described above can be written as: $f(d,\theta) = \theta_0 + \theta_1 f^0(d,\theta^*)$ where $f^0(d,\theta^*)$ represents the standardized model parameterized by the vector θ^*, θ_0 a location parameter, and θ_1 a scale parameter.[3] For example, the standardized model for Emax model can be written as $f(d,\theta^*) = \frac{d}{d+ED_{50}}$, where $\theta^* = ED_{50}$.

Figure 3 is the plot generated for all explained models. In this plot, parameters of each model have been assumed and the dose-response models are generated using the *Mods* function of package *DoseFinding* [14] .

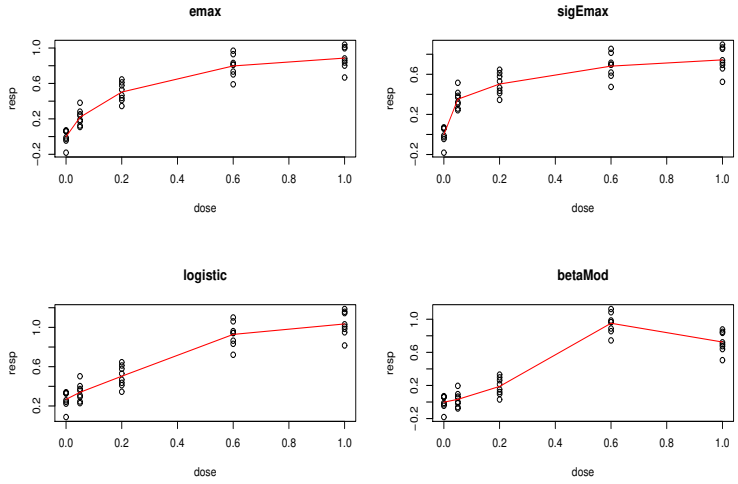

Figure 4: Simulated dose-response data for different models with values at doses = 0, 0.05, 0.2, 0.6, 1, Sample size = 8 for each dose, Between subject standard deviation for each dose is 0.1, Model parameters for all standardized models are, for Emax: ED_{50} =.2 , For SigEmax: ED_{50} =.2 and N = 0.5, For logistic: ED_{50} = 0.2 and δ = 0.2, For Beta model: δ_1 = 2, δ_2 = 1.2, and scal = 1.2

2.2 Data Simulation

The general framework is that a response Y (which can be an efficacy or a safety endpoint) is observed for a given set of parallel groups of patients corresponding to doses d_2,d_3.....d_k and control d_1, for a total of k arms. For the purpose of testing PoC and estimating target dose, we consider a one-way layout for the model specification[15]

$Y_{ij} = \mu_{d_i} + \epsilon_{ij}$, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and i = 1,2,..,k arms and j = 1,2,....,n_i

where the mean response at dose d_i can be represented as μ_{d_i} = f(d_i,θ) for some dose-response model f(.) parameterized by a vector of parameters θ, n_i is the number of patients allocated to dose d_i, and ϵ_{ij} is the error term for patient j within dose group i and these are independent across subjects and also across measurements with in each subject. σ is considered as standard deviation between subjects in a particular dose. It is to be noted that, we have considered the similar variance in the observations at each dose, in practise we use the log transformed response at each dose which in turn reduce the difference between the variance in each dose, so it is safe to use the same variance across each dose while doing simulations.

To simulate the data set for a particular model, we require the doses, number of samples required for each dose, between subject standard deviation for each dose , and parameters for the model. We decided the required values based on the designed experiment and simulated specific dose-response model using *genDFdata* function of package *McpMod* [16]. This function simulates normally distributed dose-response data, according to a pre-specified dose-response model and a given standard deviation.

Figure 4 shows the simulated data for E_{max}, $SigE_{max}$, logistic, and beta models for assumed values.

Depending on the designed experiment, we have generated different data sets using different models. As we have discussed, different model shape represent different kind of dose-response relationship. To create a data set which has observations from different underline dose-response relationship, we have simulated data from different models and merged them to create the complete data set.

3 Theory

3.1 Multiple Comparison Procedure(MCP)

Generally, two types of errors can occur when conducting a test, either the null hypothesis is rejected when it is true (type-I error) or it cannot be rejected although it is not true (type-II error). Formally, the type-I error (denoted by α) is defined as

α = P(H0 rejected | H0 true)

One of the main issues in testing is to keep the type-I error below a certain designated level which is referred to as the significance level, usually of a value of 5% .

In the framework of dose-finding, it is usually the case that more than one (pairwise) comparison has to be drawn among the different dose groups. Each of those k comparisons is represented by one null hypothesis H_0^i ; i = 1,.., k. Conducting each of the pairwise comparison tests at the same (local) level α can eventually produce a rate of false positives (meaning erroneously rejected null hypotheses) above this predefined significance level. However, different definitions of error rates for a multiple testing procedure are in place.

The most conservative one is the Family Wise Error Rate (FWER), defined as the probability of committing at least one type-I error, e.g. the probability to erroneously reject any of the k null hypotheses in the whole set of comparisons:

FWER = P(number of false positives $>$ 0)

Generally, there are four main types of MCPs[17]:
1. All-contrast comparisons: all contrasts are to be tested
2. All-pairwise comparisons: all pairwise differences are to be tested
3. Multiple comparisons with the best: all treatment/dose groups shall be tested against the treatment/ dose with the best effect
4. Multiple comparisons with the control: all treatment/dose groups shall be tested against the control/active control group control.

The last type of Multiple Comparison Procedure were traditionally considered in dose-finding studies when the primary objective was to conclude whether one, or possibly more than one, dose was significantly superior than the control with respect to the primary efficacy criterion. Under the MCP approach, one evaluates the statistical significance of contrasts between doses (typically active dose vs control), while preserving the family-wise error rate (FWER) at a pre-specified level α. Proof of Concept(PoC) is established when at least one contrast is statistically significant, in which case one selects the minimum effective dose (MED) that is statistically superior to control and produces a clinically relevant effect.[18]

A large number of multiple comparison procedures are available : Few of them are methods based on ordered p-values(Bonferroni [19], Holm [20]), closed testing procedures (Hommel[21] and Hochberg [22]), and multiple contrast test (Tukey[23] Dunnett [24]).

We will discuss the multiple contrast test in more detail as this is widely used in dose-response studies.

3.1.1 Multiple Contrast Test(MCT)

The principle of (multiple) contrast tests allows testing of a more general set of hypotheses. Not only pairwise comparisons of an active dose and the control can be addressed, but all types of comparisons listed in above section. In this method, the null hypotheses are no longer formulated directly on the basis of the group means themselves but by means of contrasts of these group

means. A contrast is a linear combination of the group means $\sum_{i=1}^{k} c_i\mu_i$ with the restriction that all c_i's sum up to 0. One could also express this by the product of $c^T\mu$ with the vector forms $\mu = (\mu_1, \mu_2, ...\mu_k)^T$ and $c = (c_1, c_2, ...c_k)^T$ of the group means and contrasts respectively. Using this we can do any planned comparison using any linear combination of group means by defining the null hypothesis as $\sum_{i=1}^{k} c_i\mu_i = 0$. For example, if want to compare two dose levels with response of group means μ_1, μ_2 the contrast coefficient will be c_1 = 1 , c_2 = -1 and Null hypothesis will become μ_1 - μ_2 = 0 or there is no difference in the response between two dose levels.

Assuming the responses Y_{ij} ; i = 1,2,..,k arms and j = 1,2,....,n_i to be normally distributed and independent within and across the different dose groups, the following test statistic can be used for testing the null hypothesis $H_0 : c^T\mu = 0$

$$T(Y) = \frac{\sum_{i=1}^{k} c_i\bar{Y}_i}{S\sqrt{\sum_{i=1}^{k} c_i^2/n_i}} \tag{1}$$

Where Y is the matrix of all responses, $\bar{Y}_i$ denotes the mean response in dose group i and $S^2 = \sum_{i=1}^{k}(Y_{ij} - \bar{Y}_i)^2/\nu$ and $\nu = \sum_{i=1}^{k} n_i - k$ degrees of freedom

As the test statistic consists of the normally distributed contrast $\sum_{i=1}^{k} c_i\bar{Y}_i$ divided by the independent chi-squared distributed estimator of the pooled variance, under the null hypothesis H_0, the test statistic follows a central t-distribution $T(Y) \sim t_\nu$ with ν defined as above. This implies a rejection of the null hypothesis if the value of the test statistic exceeds the (1 - α) quantile of the central t-distribution $t_{1-\alpha,\nu}$ in the case of a single one-sided contrast test.

In case of an MCT, when testing several contrasts simultaneously, if m contrasts are to be tested, one could use the (1 - α/m)-quantile of the central t-distribution instead of the (1 - α) -quantile for each individual (one-sided) test in order to adjust according to the Bonferroni method [25]. This, however, leads to a rather conservative control of the FWER.

By defining the null hypotheses by means of contrasts, e.g. $H_0 : c^T\mu = 0$ as mentioned above, it is possible to address every hypothesis as long as only linear components of the means are involved. Particularly, by specifying appropriate contrasts, every set of pairwise comparisons can be defined. This also includes the four different types of MCTs presented in previous section. Depending on the type of MCT, the contrast vectors for all pairwise comparisons that are involved in this particular contrast test are set up in a matrix, called contrast matrix:

$$\mathrm{C} = \begin{pmatrix} \mathrm{c}_1^T \\ .. \\ \mathrm{c}_i^T \\ .. \\ \mathrm{c}_m^T \end{pmatrix}$$

Where with the single contrasts $c_i^T = (c_i^1, ..., c_i^k)$ as row vectors

By multiplying this matrix with the vector of mean responses, one obtains a vector of all pairwise differences that are to be tested. Hence, also the corresponding null hypothesis can be expressed by means of this matrix: $H_0 : C\mu = 0$. It is rejected if the maximum of the individual test statistics constructed according to the 1 exceeds the quantile of a m-dimensional t-distribution $t^m_{1-\alpha/2,\nu,R}$ where m is the number of (pairwise) hypotheses that are part of the test and R is the matrix of correlation between contrast test statistics and can be written as follows[15]: $R = \rho_{ij} = \frac{\sum_{l=1}^{k} c_{il}c_{jl}/n_l}{\sqrt{(\sum_{l=1}^{k} c_{il}^2/n_l)(\sum_{l=1}^{k} c_{jl}^2/n_l)}}; 1 \leq i, j \leq m$

One example of MCT is Dunnett procedure [24] is a method for the comparison of multiple treatment groups with a control. In this a contrast matrix is defined only containing the contrasts corresponding to the comparisons of the active treatment groups with the control group.

3.2 Modeling Procedure(Mod)

The model-based approach assumes a functional relationship between the response and the dose, taken as a quantitative factor, according to a pre-specified parametric model, such as a logistic or an Emax model. The fitted model is then used to test if a dose–response relationship is present (PoC) and, if so, to estimate an adequate dose(s) to achieve a desired response using, for example, inverse regression techniques.

Response Y (which can be an efficacy or a safety endpoint) is observed for a given set of parallel groups of patients corresponding to doses d_2,d_3.....d_k and control d_1, for a total of k arms and can be written as follows: $Y_{ij} = f(d_i, \theta) + \epsilon_{ij}$, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and i = 1,2,..,k arms and j = 1,2,....,n_i, where the dose–response model f (.) may be a linear or nonlinear function of the parameters θ.

The testing of PoC and the selection of target doses using a modeling approach requires the estimation of the model parameters θ. Under the assumption of independent, identically distributed errors ϵ_{ij} adopted here, ordinary least squares (OLS) estimates that minimize the residual sum of squares $\sum_{i=1}^{k} \sum_{j=1}^{n_i} |Y_{ij} - f(d_i, \theta)|^2$ are typically used. In the case of nonlinear dose–response models, nonlinear least squares algorithms are needed to estimate θ. The most popular of these is the Gauss-Newton algorithm [26] which is an iterative procedure consisting of solving, until convergence, a sequence of linear least squares problems based on a local approximation of the nonlinear model. Such iterative algorithms typically require a starting point, the so-called initial values, for the parameters.

The Gauss-Newton algorithm for nonlinear least squares is implemented in mainstream statistical software packages. The functions nls[27] and gnls [28] are implemented in R.

The application of a model-based approach to analyze a given dose–response study may raise questions on the validity of the statistical results, since the true underlying dose–response relationship under investigation is typically unknown. This model uncertainty remains during the entire drug development until the late Phase III clinical trials.[29]

3.3 MCPMod Method

MCP approach makes the inherent assumption that one of the doses under investigation is the correct one. In particular, the objective of finding individual doses that achieve statistical significance has been an important driver toward dosing designs with few doses. Modeling approach provides greater flexibility than the MCP methodology; however, its validity depends on an appropriate dose-response model being prespecified for the analysis. Because the model prespecification needs to be included in the study protocol, prior to any data being available, considerable prior knowledge about the dose-response shape is required for this approach to be reliable.[15]

We can say modeling techniques provide a flexible tool for describing functional dose–response relationships and subsequently selecting a suitable dose. Typical model-based analyses, however, do not provide a rigid error control, as it is provided, for example, by multiple comparison procedures. There is a hybrid approach which provides the flexibility advantages of modeling based techniques within the framework of multiple comparisons abbreviated MCPMod.[3] The motivation for MCP-Mod is derived from the recognition that the power of standard dose-response trend tests depends on the (unknown) dose-response relationship and proposed the use of a set of multiplicity adjusted trend tests based on several transformations of the doses.[30]

In this method, we assume the existence of several candidate parametric models and use multiple comparison techniques to choose the one most likely to represent the underlying dose-response curve. The first step consists of computing optimum contrast coefficients, once the doses, sample sizes, and the set of candidate models have been identified. In the second step we evaluate the significance of the individual models in terms of the corresponding single contrast tests. Good candidate models are chosen which lead to significant contrast tests. Such a procedure allows the

selection of the most adequate dose-response model, while preserving the FWER. Other criteria may be used for the model selection step, which is also discussed briefly. Finally, the selected model is then used to produce inference on adequate doses, employing a model-based approach.

Formulation of MCPMod method:[15]

As described in section 2.2, the general framework is that a response Y (which can be an efficacy or a safety endpoint) is observed for a given set of parallel groups of subjects corresponding to doses d_2,d_3.....d_k and control d_1, for a total of k arms. For the purpose of testing PoC and estimating target doses, we consider a one-way layout for the model specification[15]

$$Y_{ij} = \mu_{d_i} + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \tag{2}$$

for i = 1,2,..,k arms and j = 1,2,....,n_i, where the mean response at dose d_i can be represented as μ_{d_i} = f(d_i,θ) for some dose-response model f(.) parameterized by a vector of parameters θ, n_i is the number of patients allocated to dose d_i, and ϵ_{ij} is the error term for patient j within dose group i.

Most dose-response models used in practice can be written as: $f(d, \theta) = \theta_0 + \theta_1 f^0(d, \theta^*)$ where $f^0(d, \theta^*)$ represents the standardized model parameterized by the vector θ^*, θ_0 a location parameter, and θ_1 a scale parameter.[3]

Assume that a set S of M parameterized candidate models is given together with prior parameter values for their standardized versions. When used with the doses planned for the trial, d_1,d_2.....d_k, these candidate models produce mean response vectors $\mu_m = (\mu_{m1}, \mu_{m2}...., \mu_{mk})$ where $\mu_{mi} = f_m(d_i, \theta_m)$, and f_m is the m_{th} dose-response function, m = 1,2,...M. Each of the dose-response shapes in the candidate set is tested using a single contrast test, with coefficients chosen to maximize the power of the test when the true underlying mean response equals μ_m.

The single contrast tests are defined as

$T_m = \frac{\sum_{i=1}^{k} c_{mi} \bar{Y}_i}{S* \sqrt{\sum_{i=1}^{k} c_{mi}^2 / n_i}}$

where $S^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2/(N-k)$ and $N = \sum_{i=1}^{k} n_i$. The contrast vectors $c_m = (c_{m1}....c_{mk})'$ follow the regularity conditions $\sum_{i=1}^{k} c_{mi} = 0$ and $\sum_{i=1}^{k} c_{mi}^2 = 1$. Optimum contrast coefficients for model testing are determined as the best contrast associated with a given model function f(d,θ), in the sense that, when that model is correct, it maximizes the chance of rejecting the associated null hypothesis.[3] The associated null hypothesis are $H_0^m : c_m^T \mu = 0$ for given k $\times$ 1 contrast vectors $c_m^T = (c_{m1}, ..., c_{mk})$ of known constants subject to $c_m^T 1 = 0$, m = 1,. . . ,M models.

Every single contrast test thus translates into a decision procedure to determine whether a given dose-response shape is statistically significant, based on the observed data. This means for a particular model the associated Null hypothesis $H_0^m : c_m^T \mu = 0$ is rejected and the observed data significantly follow the underlying dose-response curve. The final test statistic T_{max} is based on the maximum contrast test, i.e., $T_{max} = max_m T_m$. PoC is verified if this maximum statistic T_{max}, and thus at least one single contrast test, is statistically significant, while controlling the FWER at level α. Let $q_{1-\alpha}$ denote the multiplicity adjusted critical value. PoC is then established if $T_{max} \geq q_{1-\alpha}$. Under the distributional assumptions stated in 2, the contrast test statistics T_1....T_M are jointly multivariate t distributed with correlation matrix R determined by the model contrasts and N - k degrees of freedom.

Numerical integration routines [31] or Monte carlo simulation can be used to compute the multiplicity adjusted p-values and critical values. If no candidate model is statistically significant, the procedure stops indicating that a dose-response relationship cannot be established from the observed data (i.e., no PoC). Otherwise, the maximum contrast test statistic, and possibly further contrasts, are statistically significant that is, larger than $q_{1-\alpha}$. Note that testing the individual contrasts at multiplicity adjusted level ensures the strong control of the FWER. For the PoC step, the same multiple comparison procedure was used to test, the global hypothesis of no dose-response,

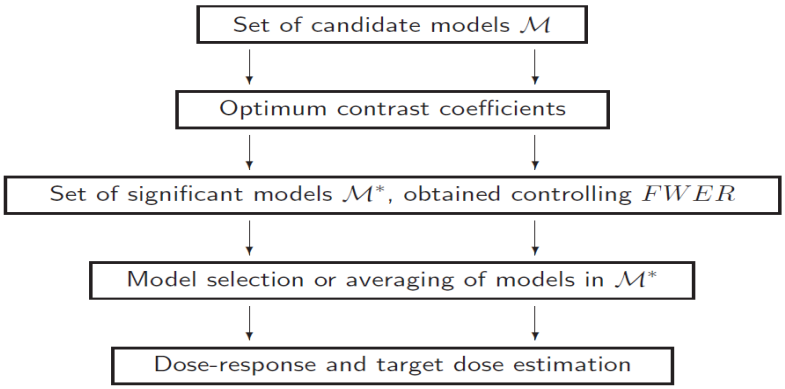

Figure 5: Overview of MCPMod method [32]

without the individual hypotheses being formally tested.

Out of the statistically significant models in the candidate set a best model is selected for dose estimation in the last stage of the procedure. The selection of the dose estimation model can be based on the minimum p-value or some other relevant model selection criteria like the AIC or the BIC.[3] The selected dose-response model is then employed to estimate target doses using inverse regression techniques and possibly incorporating information on clinically relevant effects. The precision of the estimated doses can be assessed using, for example, bootstrap methods. The complexity of implementing such dose estimation procedure will vary with application and can be computationally involved in some cases.

Figure 5 illustrates schematic overview of the multiple comparison and modeling procedure.

4 Methodology and Designed Experiments

4.1 Methodology

For all the experimental study, methodology can be divided in two parts: Design and Analysis.
We describe below the steps for both parts while using the MCPMod method. To apply the methodology we have used functions of the packages *DoseFinding* [14] and *McpMod* [16] in R[13].

4.1.1 Design

After defining the main study characteristics (e.g. the primary endpoint and the study population) as it is essential for any study, the study-specific features have to be set up such that the outcome of the study is as promising as possible. Those features include[25]:

1. candidate models for the dose-response relationship (selected on the basis of available prior knowledge which can be obtained for example from similar compounds),

2. the choice of dose groups to be included in the study as well as the corresponding allocation ratios,

3. the optimal contrasts for the selected candidate models to maximize the power of the trend tests conducted in the MCP-step,

4. the sample size providing a certain target power for the establishment of PoC or a certain precision for one of the estimates of interest.

Optimal design criteria
We can search for the optimal selection of dose groups to be included in the study and identify how to allocate the animals optimally to the selected dose groups. The ratio of required animals at each dose level is referred to as an allocation ratio. The identification of optimal design features is implemented in the function *optDesign* of *DoseFinding* [14] package in R , which offers three different kinds of optimality criteria. Either the study design is optimized with regard to the estimation of the model parameters (D-optimal), with regard to the Target Dose(TD) estimation (TD-Optimal) or with regard to both. In practice, D-optimal signifies the minimization of a criterion which involves the variance of the model parameters whereas TD Optimal means minimizing the length of the confidence interval for the TD.

4.1.2 Analysis

Once the study has been designed thoroughly and data is simulated accordingly, the analysis is carried out in two subsequent steps. First, the models are tested separately for an existing dose effect while still adhering to the overall type-I error. If that could be proven for at least one of the candidate models, the dose-response relationship is modelled by means of a parametric model that is either the candidate model that fits the data best or a weighted average over those candidate models with a significant test result. On the basis of the fitted model, the target dose can be estimated via inverse regression techniques and precision of the estimates can be assessed if desired.

In the dose finding step there are two methods using which we can estimate the dose (i) using best fitted model (ii) using weighted average of all fitted models. To use the first method we need to select the best fitted model. The best fit model can be selected with the smallest p-value or highest T-value as it is most likely to be (closest to) the true model or the best model with respect to some goodness-of-fit criterion. As mentioned in section 1.2 it is better to use a criterion like AIC. In this thesis we will use AIC criterion to decide the best fit model and then we can estimate the dose by using best fit model.

To use the second method we use the posterior weight of all fitted models. This can be inferred as how much weight to the particular models should be given if we want to use all the fitted models

for inference. These weights for all fitted models can be calculated using the equation:

$$w_i = \frac{\exp\left(0.5 \cdot AIC_i\right)}{\sum_{j=1}^{k} \exp\left(0.5 \cdot AIC_j\right)} \tag{3}$$

where w_i is the posterior weight of i^{th} fitted model, AIC_i is AIC value for i^{th} fitted model , and k is number of all fitted models. After calculating the posterior weight we can calculate the estimated dose by using weighted average of all fitted models, which means Estimated dose(ED) = $\sum_{i=1}^{k} w_i \cdot ED_i$ where w_i is the posterior weight of i^{th} fitted model and ED_i is estimated dose calculated by i^{th} fitted model.[33]

4.2 Designed Experiments

MCPMod method can be used in clinical trials to simultaneously address the both aims of verifying PoC and estimating the dose.[15] In this thesis we want to investigate the use of the MCPMod method for pre-clinical studies.

In pre-clinical experiments, we follow the 3Rs(Replacement, Reduction, Refinement)[34] principle. The reduction principle suggests that we need to design the experiment in such a way that we require as few animals as possible. As compared to clinical trials, in these experiments the variance between the subjects in response is also smaller as compared to clinical trials. This is due to the fact that animals are genetically more similar and also in a more controlled external environment. We attempt to design the experiment in such a way that we need smaller sample sizes and generate data with low variance so that it is similar to a pre-clinical setup.

Before designing the experiment it is important to define and understand the objective of the experiments. In this thesis we want to evaluate the MCPMod method for Pre-clinical Experiment and attempt to answer the following questions:

1. Is there any reduction in sample size required for the experiment when using different allocation ratios for different dose levels(optimal design) as compared to using similar allocation ratios for each dose level(normal design)? [For results refer to 4.2.4]

2. Can we use the MCPMod method to estimate the underlying model of given data? How close are the estimates of the fitted model to the real model? While estimating the parameter of the best fitted model, what is the difference in the results using optimal design and normal design? [For results refer to 5.1]

3. What is the probability of estimating the best fit model same as the model from which data has been generated? In the Mod step of MCPMod, can we use the weighted average of all fitted models instead of the best fit model based on AIC values? [For results refer to 5.2]

4. Using MCPMod, can we detect that there are toxic effects in the responses at higher doses? [For results refer to 5.3]

The answer to the first question can be found while designing the study for different experiments. To answer the second question we will design two experiments, one with optimal design (different allocation ratio for different dose levels) and another with normal design (equal allocation ratio for each dose level). To answer the third and the fourth question we will use the optimal design and attempt to answer these by analysing data generated for 2000 simulations.

Design:

4.2.1 Candidate Model Selection

For both the experiments with optimal design and normal design, we use the same candidate models.

After discussing with the modeling team we have selected six candidate models: Three Emax models, one Sigmoid Emax, one logistic model, and one Beta model. We have selected different models so that we can cover different shapes of the dose-response relationship. The reason for selecting three Emax models is that it is the most common found dose-response model but we can not be sure about the parameters of the model in advance so it is advisable to use several models with different parameters. The beta model can be used to detect the toxicity at higher doses.

We find the parameters of the standardized version of each model which is described in Section 3.3 using function *guesst* of package *DoseFinding*. [14].

To use this function, we need prior information about the expected dose-response relationship. We need to know the number of (dose, response) pair values same as the number of unknown parameters in the standardized chosen models. In our case for Emax we need one (dose, response) pair, for Betamod we need one (dose, response) pair and the value of the dose at which the maximum effect occurs, and for Sigmoid Emax and logistic models we need two (dose, response) pair values. Criteria to choose these values are generally based on previous knowledge or historical data. In our case, we assume that we have some idea of the expected dose-response relationship and that we can guess the value at dose = 0.1 and dose = 0.5 which are 10% and 50-80% of the maximum effect respectively. For the Sigmoid Emax and logistic models which need two pairs of values, we will be using response values 10% and 50% at dose = 0.1 and 0.5 respectively. For BetaMod we will use response value 70% at dose = 0.5 and maximum effect occurs at dose = 0.75. For one of the Emax model we use response value 70% at dose = 0.5 and for the other two Emax models we will use response values 60% and 80% at dose = 0.5. This way we have ensured that we have 3 different Emax models with different parameters. Using these values we have tried to cover the different shapes of initial candidate models along with one close to the one with which we will be simulating the data.

Table 1 describes the guessed values and calculated parameters for all standardized versions of all the candidate models. Referring to figure 6 we can see that each of the Emax models have different value at dose 0.5, which is the assumed values at dose 0.5 and the model must pass through that point.

Model Name	Guessed Value	Output parameters
Emax1	per = 0.8 at d = 0.5	ED_{50} = 0.125
Emax2	per = 0.7 at d = 0.5	ED_{50} = 0.214
Emax3	per = 0.6 at d = 0.5	ED_{50} = 0.333
BetaMod	per = 0.7 at d = 0.5, Maximum Effect at d = 0.75, scal= 1.2	δ_1 = 2.54 and δ_1 = 1.52
Sigmoid Emax	per = 0.5 at d = 0.5 , per = 0.1 at d = 0.1	ED_{50} = 0.5 and N = 1.37
Logistic	per = 0.5 at d = 0.5 , per = 0.1 at d = 0.1	ED_{50} = 0.5 and $\delta = 0.182$.

Table 1: Parameters of standardized candidate models and the guessed values to calculate those(per is fraction of maximum effect and d is dose level)

We have set the value of control effect to 0 and maximum change from base effect to 1 and used the *Mods* function of the *DoseFinding* package to calculate the complete model involving all the parameters.

All complete candidate models are as follows:
First Emax model is $f(d,\theta) = 0 + \frac{d \cdot 1.125}{d+0.125}$
Second Emax model is $f(d,\theta) = 0 + \frac{d \cdot 1.214}{d+0.214}$

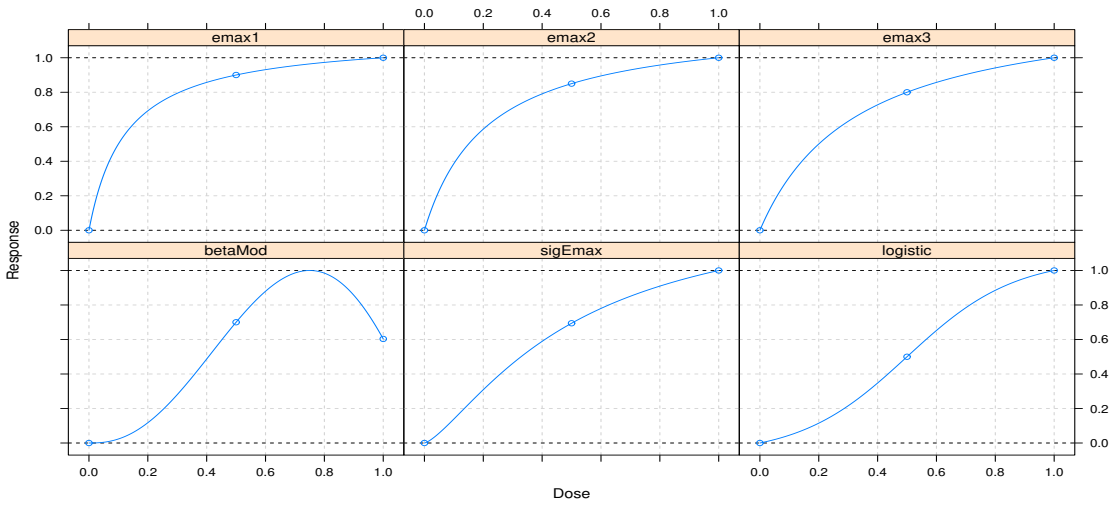

Figure 6: All candidate models. Dots on the curve are the response at dose 0,0.5,1

Third Emax model is $f(d,\theta) = 0 + \frac{d \cdot 1.333}{d+0.333}$
BetaMod Model is $f(d,\theta) = 0 + 1 \cdot B(2.54, 1.52) \cdot (\frac{d}{1.2})^{2.54} \cdot (1 - \frac{d}{1.2})^{1.52}$
Sigmoid Emax model is $f(d,\theta) = 0 + \frac{d^{1.37} \cdot 1.39}{d^{1.37} + 0.5^{1.37}}$
logistic model is $f(d,\theta) = -0.069 + \frac{1.137}{1+\exp\frac{0.5-d}{0.182}}$

Figure 6 shows the plot for all above mentioned candidate models. We can observe that different candidate models are representing different shapes for the dose-response relationship.

4.2.2 Dose levels selection and respective allocation ratio

For optimal design, to find the optimal doses and respective allocation ratios we have used the function *optDesign* of the *DoseFinding* [14] package . Our main aim is to understand the shape of the model in observed data so we have used the D-optimal optimality criteria as mentioned in section 4.1.1. The probabilities of all candidate models are assumed equal and the possible input dose levels are [0,0.05,0.1,0.15,0.2,0.25,0.4,0.45,0.5,0.6,0.7,0.8,0.9,1].

The result of optimal design is as follows:

```
Calculated D - optimal design:
      0     0.1    0.15    0.35     0.7       1
0.27075 0.07250 0.11346 0.15760 0.14712 0.23857
```

The output of this function gave us that required dose levels are [0, 0.1,0.15,0.35,0.7,1] and the ratios of sample size required for each dose level are [0.271,0.073,0.114,0.158,0.147,0.239].

To find the integer allocation ratio at different dose levels we have divided the output of the ratio at each dose level by 0.025 and then taken the nearest integer value as allocation ratio. We have divided by 0.025 as we can safely assume that 2.5% is a very low weight as compared to the total weight and we can ignore all the weights below this value. We got the allocation ratios as [11,3,5,6,6,10] for the dose levels [0, 0.1,0.15,0.35,0.7,1]. This means that if we require 11 samples at dose 0, we will require 3 samples at dose 0.1, 5 samples at dose 0.15, 6 samples at doses 0.35 and 0.7, and 10 samples at dose 1.

For the normal design, we will use the same dose levels [0,0.1,0.15,0.35,0.7,1] with allocation ratio [1,1,1,1,1].

4.2.3 Optimal contrasts for the selected candidate models

We have used function *optContr* for calculating the optimal contrast for all the candidate models. This function takes the input as candidate models and weight of each dose group or allocation ratio.

For optimal design with different allocation ratio for each dose, we got the following optimal contrasts:

```
Optimal contrasts
      emax1  emax2  emax3 betaMod sigEmax logistic
0    -0.831 -0.795 -0.761  -0.628  -0.672   -0.563
0.1  -0.043 -0.072 -0.091  -0.160  -0.131   -0.137
0.15 -0.002 -0.050 -0.085  -0.237  -0.163   -0.208
0.35  0.156  0.130  0.101   0.023   0.034   -0.100
0.7   0.249  0.262  0.268   0.605   0.279    0.278
1     0.470  0.524  0.568   0.396   0.653    0.731
```

In the above table, every column represents the contrast coefficient for each candidate model with respect to each dose. For ex. for the Emax1 model contrast coefficients are c_0 = -0.831 , $c_{0.1}$ = -0.043 , $c_{0.15}$ = -0.002 , $c_{0.35}$ = 0.156 , $c_{0.70}$ = 0.249 , c_1 = 0.470 and the null hypothesis for this model will be for this model H_0^{emax1} : $c_0\mu_0 + c_{0.5}\mu_{0.1} + c_{0.15}\mu_{0.15} + c_{0.35}\mu_{0.35} + c_{0.7}\mu_{0.7} + c_1\mu_1 = 0$ and will represent no dose-response effect for the Emax1 model, where μ_{dose_i} is mean response at $dose_i$.

For normal design with similar allocation ratio for each dose, we got the following optimal contrasts:

```
Optimal contrasts
      emax1  emax2  emax3 betaMod sigEmax logistic
0    -0.780 -0.707 -0.648  -0.387  -0.505   -0.383
0.1  -0.180 -0.248 -0.286  -0.361  -0.352   -0.335
0.15 -0.043 -0.113 -0.161  -0.318  -0.257   -0.302
0.35  0.215  0.188  0.155   0.043   0.079   -0.090
0.7   0.366  0.398  0.413   0.728   0.435    0.442
1     0.421  0.482  0.527   0.295   0.600    0.668
```

In the above table, every column represents the contrast coefficient for each candidate model with respect to each dose. For ex. for the Emax1 model contrast coefficients are c_0 = -0.780 , $c_{0.1}$ = -0.180, $c_{0.15}$ = -0.043 , $c_{0.35}$ = 0.215 , $c_{0.70}$ = 0.366 , c_1 = 0.421 and the null hypothesis for this model will be for this model H_0^{emax1} : $c_0\mu_0 + c_{0.5}\mu_{0.1} + c_{0.15}\mu_{0.15} + c_{0.35}\mu_{0.35} + c_{0.7}\mu_{0.7} + c_1\mu_1 = 0$ and will represent no dose-response effect for the Emax1 model, where μ_{dose_i} is mean response at $dose_i$.

4.2.4 Sample size calculation

We have used the *sampSizeMCT* function of the *DoseFinding* package to calculate the sample size required for each dose level. To calculate these we have assumed significance level $\alpha = 0.05$ for a one-sided hypothesis test, power to detect PoC as 80%, and expected standard deviation with in subjects at each dose level as 0.75 for both the optimal and the normal design experiments. We have chosen these values through consultation with experts.

For the optimal design experiments with different allocation ratios for each dose level, we have used the samples allocation ratios calculated in step 2 which are [11,3,5,6,6,10] for dose levels [0,0.1,0.15,0.35,0.7,1], the optimal contrasts calculated in step 3 and got that we need sample sizes [7,2,3,4,4,7] for dose levels [0,0.1,0.15,0.35,0.7,1] respectively.

For the normal design experiments with different allocation ratios for each dose level, we have used the samples allocation ratios calculated in step 2 which are [1,1,1,1,1,1] for dose levels

[0,0.1,0.15,0.35,0.7,1], the optimal contrasts calculated in step 3 and got that we need sample sizes [6,6,6,6,6,6] for dose levels [0,0.1,0.15,0.35,0.7,1] respectively.

Discussion on the sample sizes required by the optimal and normal designs:
The total sample size required with optimal design is 27 samples, while with normal design a total 36 samples are required. We can say that in this case we can achieve a 25% reduction of the total sample size required. In turn this will also reduce the total dose required in the experiment. It is to be noted that this value depends on the design criteria, input parameters such as selected candidate models, power to detect PoC value, significance level, and the expected between subject standard deviation.

In addition to the reduction in the total sample size required, another advantage of using the optimal design is that we were able to filter out the doses levels we do not require.

At this point we can say that there is an improvement in the total sample size required when using optimal design as compared to normal design, however we still need to check how this effects the result of the experiments. Going forward we will simulate the data for both designed experiments and evaluate the results.

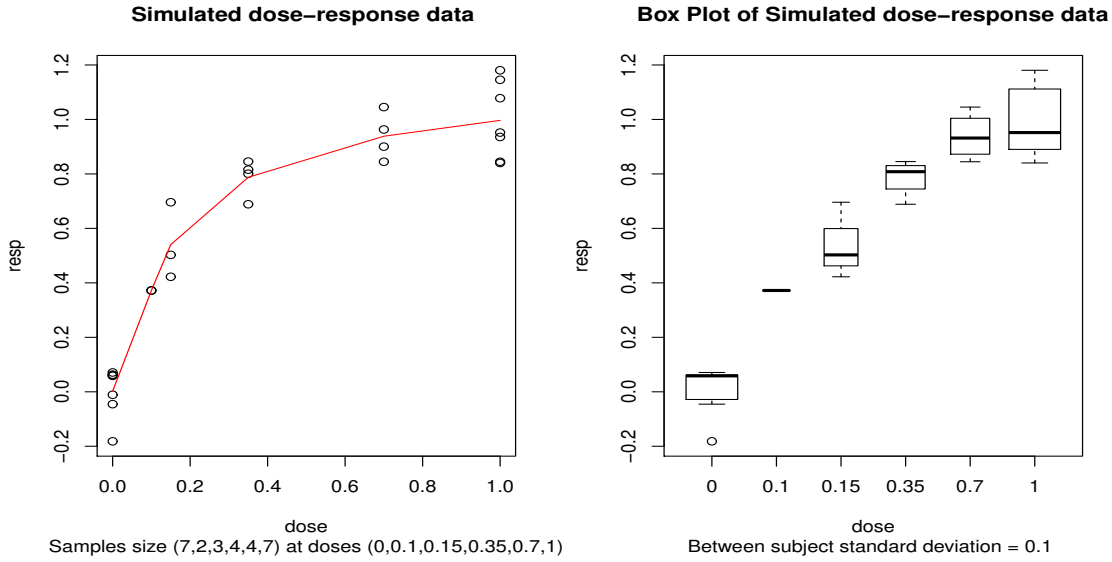

Figure 7: Simulated dose-response data from the Emax model with parameters $E_0 = 0$, $E_{max} = 1.2$, $ED_{50} = 0.2$, dose levels = [0,0.1,0.15,0.35,0.7,1] , different allocation sample sizes = [7,2,3,4,4,7] and between subject standard deviation = 0.1 , red line is joining the means at each dose

5 Simulations and Results

5.1 Evaluation of MCPMod for Pre-clinical Experiments

We have generated two data sets for dose levels [0,0.1,0.15,0.35,0.7,1] using both optimal design and normal design. For this, we have used the Emax model with parameters $E_0 = 0$, $E_{max} = 1.2$, $ED_{50} = 0.2$,and with the following setup:
1. Sample sizes [7,2,3,4,4,7] and between subject standard deviation = 0.1 at each dose level
2. Sample sizes [6,6,6,6,6,6] and between subject standard deviation = 0.1 at each dose level

In the above setup sample sizes are decided based on the sample size calculated in the design step, and between subject standard deviation is decided based on the advice from experts with knowledge of pre-clinical experiments historical data.

Simulated data points and box plots at each dose for both data sets are shown in Figure 7 and Figure 8 respectively. Full data is presented in A.1. We can observe that in both data sets means at each dose are following the dose-response relationship of the Emax model.

Assuming maximum effect to be 1, we can calculate the true value of estimated dose for 50% effect of the maximum effect by using the definition of the Emax model and known parameters of the underlying data generating model. This results in an Estimated Dose(ED) of 0.1429.

We have applied the MCPMod methodology on both above simulated data sets and run the analysis using the *MCPMod* function of package *DoseFinding*. [14]

We have conducted one-sided hypothesis testing at significance level= 0.05, and selected the best fit model based on AIC values.

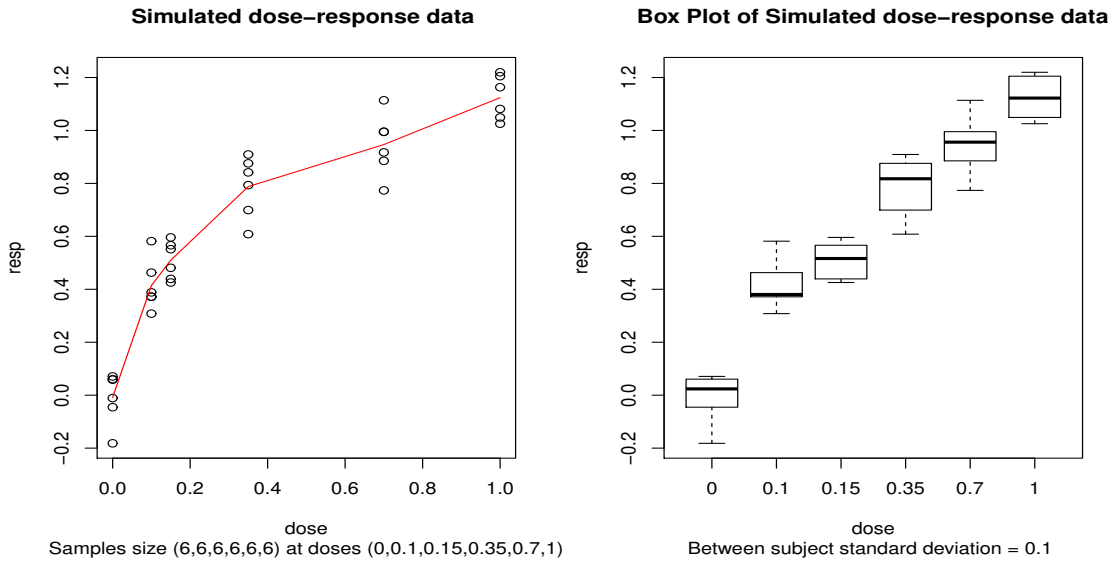

Figure 8: Simulated dose-response data from the Emax model with parameters $E_0 = 0$, $E_{max} = 1.2$, $ED_{50} = 0.2$, dose levels = [0,0.1,0.15,0.35,0.7,1] , similar allocation sample sizes = [6,6,6,6,6,6] and between subject standard deviation = 0.1, red line is joining the means at each dose

For the first data set with optimal design and sample sizes [7,2,3,4,4,7], we have got the following results(Summary of the full R output is presented at A.2):

In the first step of the MCPMod method(MCP), all the selected candidate models were significant and we can see for all the models the multiplicity adjusted p-values are less than 0.001 and the corresponding t-statistic values are significantly higher than the t-critical value of 2.032.

```
Multiple Contrast Test:
         t-Stat  adj-p
emax2    19.289 <0.001
emax1    19.223 <0.001
emax3    19.164 <0.001
sigEmax  18.393 <0.001
logistic 16.850 <0.001
betaMod  16.426 <0.001

Critical value: 2.032 (alpha = 0.05, one-sided)
```

AIC values for all fitted models are given below and the minimum AIC value is achieved for Emax model so Emax is the best fitted model based on AIC values.

```
Model selection criteria (AIC):
****************************************
     emax   betaMod   sigEmax  logistic
-42.01810 -38.81453 -40.33912 -37.95860

Selected model: emax
```

Table 2 describes the estimates, standard errors, and 95% confidence intervals for all estimated parameters for all the fitted models and Figure 9 shows the fitted model with 95% confidence intervals. We can observe that all the models fit the data and are significant but Emax seems to be the best fit as suggested by AIC values. Comparing the estimated parameter(E_0 = 0, E_{max} = 1.19, ED_{50} = 0.188) values of the fitted Emax model with the model from which the data has been simulated (E_0 = 0, E_{max} = 1.2, ED_{50} = 0.2) we can say that the estimates are good and the method is able to detect the underlying model.

Model name	Estimated Parameters	Standard Error	95% CI for Parameters
emax	e0 = 0.00	0.0382	e0: [-0.075,0.075]
	eMax = 1.19	0.0748	eMax : [1.043,1.337]
	ed50 = 0.188	0.0440	ed50: [0.102,0.275]
sigEmax	e0 = 0.002	0.0390	e0: [-0.075,0.078]
	eMax = 1.195545567	0.1441	eMax: [0.814,1.379]
	ed50 = 0.162	0.0419	ed50: [0.080,0.245]
	N = 1.235	0.4515	N: [0.35,2.12]
logistic	e0 = -0.942	1.5521	e0: [-3.984,2.10]
	eMax = 1.914	1.5673	eMax: [-1.158,4.986]
	ed50 = 0.001	0.2153	ed50: [-0.421,0.422]
	δ = 0.135	0.0619	δ: [0.014,0.257]
betaMod	e0 = 0	0.0401	e0: [-0.078,0.079]
	eMax = 1.005	0.0517	eMax: [0.904,1.106]
	δ_1 = 0.479	0.1060	δ_1: [0.271,0.687]
	δ_2 = 0.159	0.0969	δ_2: [-0.031,0.349]

Table 2: Parameters for all fitted models along with their 95% CI (Used data set with Sample sizes [7,2,3,4,4,7])

In the second step of the MCPMod method, we have estimated the dose required to have 50% of the maximum observed effect. We have estimated this value with both the methods using the best fitted model and using weighted averages of all the fitted models. Below is the dose required using all the fitted models and posterior weights(calculated using equation 3) of all fitted candidate models:

```
Estimated ED, p=0.5
************************************
    emax  betaMod  sigEmax logistic
  0.1368   0.1389   0.1390   0.1489
Model weights (AIC):
    emax  betaMod  sigEmax logistic
  0.5666   0.1142   0.2447   0.0744
```

Estimated dose required using the best fitted model(emax) = 0.1368
Estimated dose required when using the weighed averages of all the fitted models = 0.1385

We can say that both of these values are close to the expected true value 0.1429. We will do more analysis on this and calculate the confidence interval for the estimated dose in the next section 5.2.

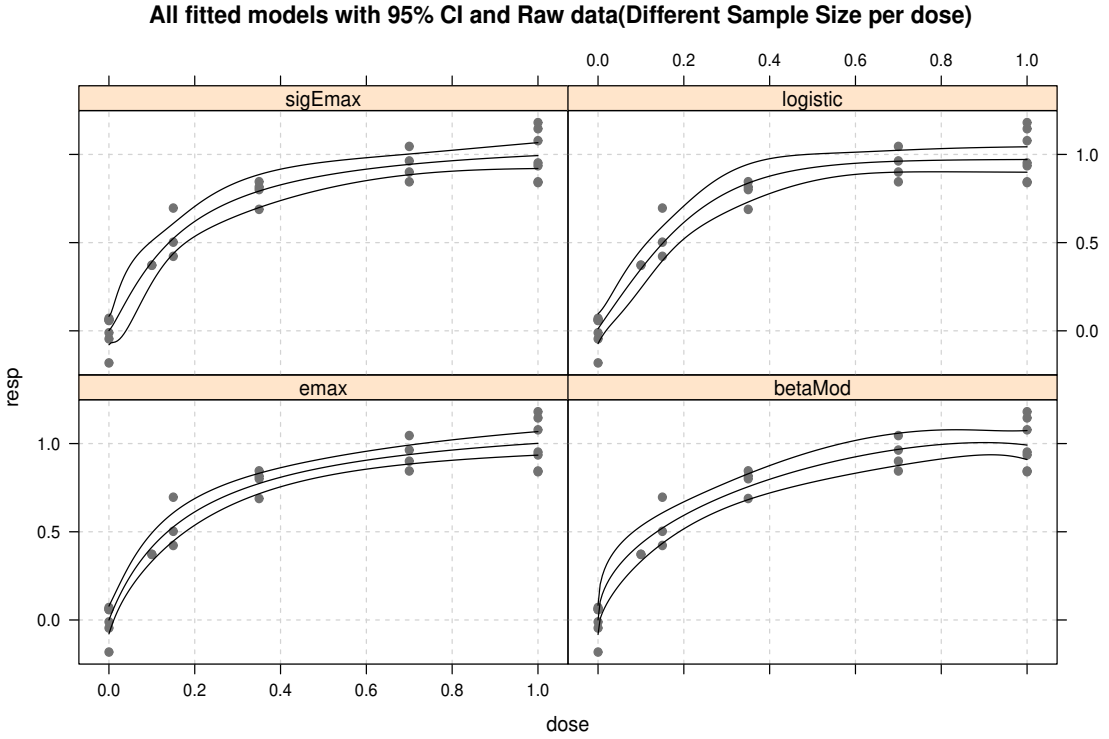

Figure 9: All fitted models along with 95% confidence intervals with all data points of data set with sample sizes [7,2,3,4,4,7] and between subject standard deviation = 0.1

For the second data set with normal design and sample sizes [6,6,6,6,6,6], we got the following results(Summary of the full R output is presented at A.2):

In the first step of the MCPMod method(Multiple comparison procedure), all the selected candidate models were significant and we can see that for all the models adjusted p-values are less than 0.001 and corresponding t-statistic values are higher than the t-critical value 2.02.

```
Multiple Contrast Test:
         t-Stat  adj-p
emax2    22.905 <0.001
emax3    22.859 <0.001
emax1    22.592 <0.001
sigEmax  21.991 <0.001
logistic 20.308 <0.001
betaMod  18.744 <0.001

Critical value: 2.02 (alpha = 0.05, one-sided)
```

AIC values for all fitted models are given below and the minimum AIC value is for the Emax model so Emax is the best fitted model based on AIC values.

```
Model selection criteria (AIC):
************************************
     emax   betaMod   sigEmax  logistic
-60.38221 -58.97291 -59.95446 -47.63632

Selected model: emax
```

In the results of this experiment we can also observe that all the models fit the data and are significant but Emax seems to be the best fit as suggested by AIC values. Figure 10 shows the fitted models with 95% confidence intervals and Table 3 describes the estimates, standard errors, and 95% confidence intervals for all estimated parameters for all the fitted models. Comparing the estimated parameter(E_0 = 0, E_{max} = 1.343, ED_{50} = 0.243) values of the fitted Emax model with the model from which the data has been simulated (E_0 = 0, E_{max} = 1.2, ED_{50} = 0.2) we can say that the estimates are good and the method is able to detect the underlying model.

Model name	Estimated Parameters	Standard Error	95% CI for Parameters
emax	e0 = 0.00	0.0391	e0: [-0.076,0.077]
	eMax = 1.343	0.0715	eMax : [1.20,1.48]
	ed50 = 0.243	0.0437	ed50: [0.157,0.329]
sigEmax	e0 = -0.009	0.0396	e0: [-0.087,0.069]
	eMax = 1.991	1.1326	eMax: [-0.229,4.211]
	ed50 = 0.702	1.2093	ed50: [-1.669,3.072]
	N = 0.667	0.2574	N: [0.163,1.172]
logistic	e0 = -0.957	1.2926	e0: [-3.491,1.576]
	eMax = 1.998	1.3139	eMax: [-0.578,4.57]
	ed50 = 0.001	0.1931	ed50: [-0.377,0.379]
	δ = 0.156	0.0591	δ: [0.040,0.272]
betaMod	e0 = 0.01	0.0401	e0: [-0.088,0.069]
	eMax = 1.127	0.0801	eMax: [0.970,1.284]
	δ_1 = 0.443	0.0720	δ_1: [0.302,0.584]
	δ_2 = 0.05	0.0717	δ_2: [-0.09,0.19]

Table 3: Parameters for all fitted models along with their 95% CI (Used data set with Sample sizes [6,6,6,6,6,6])

In the second step of the MCPMod method, we have estimated the dose required to have 50% of the maximum observed effect. We have estimated this value with both the methods using the best fitted model and using the weighted average of all the fitted models. Below is the dose required using all the fitted models and posterior weights(calculated using equation 3) of all the fitted candidate models:

```
Estimated ED, p=0.5
****************************************
    emax  betaMod  sigEmax logistic
  0.1635   0.1773   0.1696   0.1710
Model weights (AIC):
    emax  betaMod  sigEmax logistic
  0.4341   0.2146   0.3505   0.0007
```

Estimated dose required using the best fitted model = 0.1635
Estimated dose required when using the weighed average of all the fitted models = 0.1686

We can say that both of these values are close to the expected true value 0.1429. We will do more analysis on this and calculate the confidence interval for estimate doses in next section 5.2

Comparison of results from optimal and normal designs:

We have observed that in both the optimal design and the normal design using the MCPMod method we are able to fit the data with the underlying model. To compare the parameters of the best fitted model using both the designs we have presented estimated values of the parameters in table 4. We have observed that in both the designs the best fitted model is the Emax model. Estimated

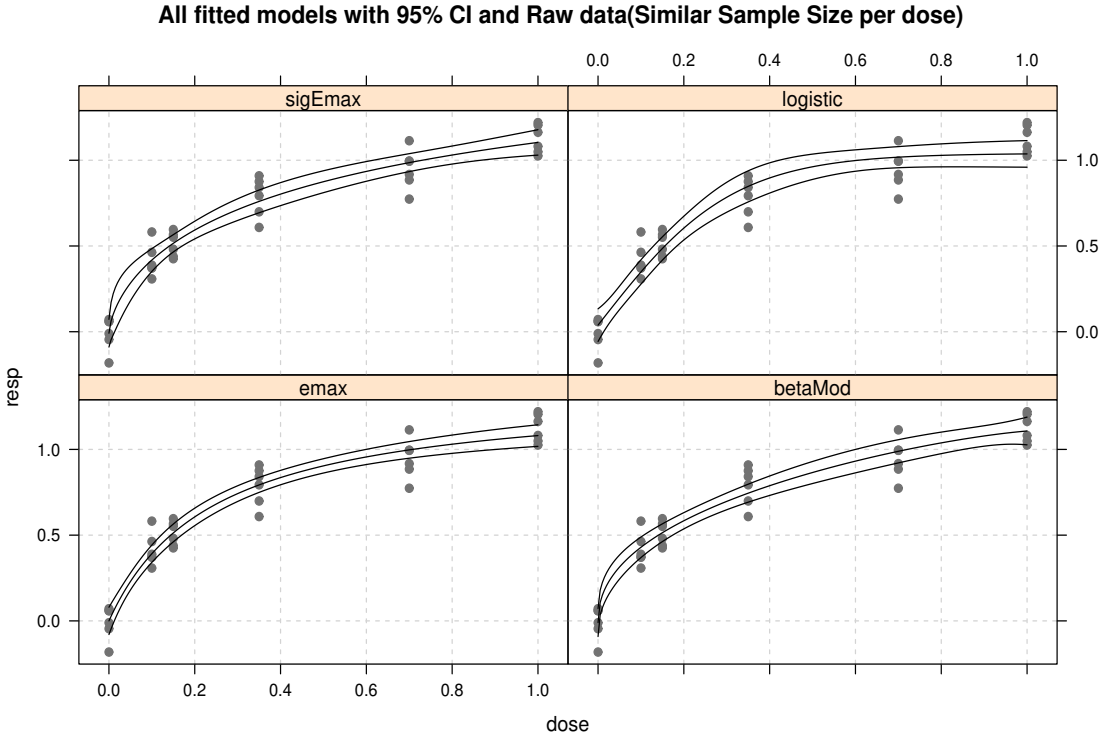

Figure 10: All fitted models along with 95% confidence intervals with all data points of data set with sample sizes [6,6,6,6,6,6] and between subject standard deviation = 0.1

parameters and residual standard errors are the same for the E_0 parameter for both designs. While for the E_{max} and ED_{50} parameters, the estimated value is higher in case of the normal design as compared to the optimal design but the standard errors are the same so the wideness of the confidence intervals are same in both cases.

Data for both the designs are different as they are generated separately and have different sample sizes so difference in the estimates is possible due to different data sets, however we can say that fitted models using both designs are fairly reasonable when compared to the model from which data has been generated which is the Emax model with parameters $E_0 = 0$, $E_{max} = 1.2$, $ED_{50} = 0.2$. By comparing the posterior weights of the best fitted model in both designs, we observe that weight of the Emax model is 0.5666 in optimal design and 0.4341 in normal design so we can say that the Emax model is better fitted with respect to its data set in optimal design.
While comparing the estimated dose at 50% of the maximum observed effect in the second step of the MCPMod method, we have already seen that results using both designs are close to the expected true value.

Best fitted Model	Estimated Parameters	Standard Error	95% CI for Parameters
emax using Optimal design	e0 = 0.00	0.0382	e0: [-0.075,0.075]
	eMax = 1.19	0.0748	eMax : [1.043,1.337]
	ed50 = 0.188	0.0440	ed50: [0.102,0.275]
emax using Normal design	e0 = 0.00	0.0391	e0: [-0.076,0.077]
	eMax = 1.343	0.0715	eMax : [1.20,1.48]
	ed50 = 0.243	0.0437	ed50: [0.157,0.329]

Table 4: Parameters for the best fitted models along with their 95% CI (For both optimal design and normal design)

5.2 Probability of selecting the true model as the best fit model and comparison of the estimated doses using the best fit model and the weighted average of all fitted models

Taking note of previous results we can say that optimal design works well so in our further experiments we will consider optimal design. In this exercise we have conducted 2000 simulations of the experiments and first calculated the probability of selecting the true model as the best fit model and then compared the estimated doses calculated using the best fit model and the weighted average of all fitted models.

We have generated data sets for 2000 simulations using Emax model with parameters E_0 = 0, E_{max} = 1.2, ED_{50} = 0.2, with the same set up as before:
Sample sizes [7,2,3,4,4,7] for doses [0,0.1,0.15,0.35,0.7,1] respectively and between subject standard deviation = 0.1 at each dose level

For every simulation we have run the analysis using the *MCPMod* function of the package *DoseFinding* [14] and estimated the best fit model based on the AIC value and the posterior weight of all fitted models and then calculated estimated doses with both the methods described in section 4.1.2.

To answer the first part of the exercise and calculate the probability of selecting a specific model as the best fit, we have calculated how many times each model has been selected as the best fitting model. Table 5 shows the number of times each model has been selected and its respective probability of selection. The probability of the selected model is calculated as the number of times the model was selected divided by the total number of simulations. We can see that Emax is selected as the best fit model in 71.6% of the simulations, much more than any of the other candidate models. It seems reasonable as data has been generated from the Emax model.

Model name	Counts of selected as Best fitted model	Percentage of selecting as best fit
Emax	1432	71.6
SigEmax	215	10.75
BetaMod	285	14.25
Logistic	68	3.4

Table 5: Models and their respective percentage of being selected as the best fit model

Table 6 shows the mean value, standard deviation, and 95% Confidence interval values for estimated doses at 50% effect of the maximum observed effect using both methods. We can see that the reported values are similar to each other and the mean values in both cases are close to the expected true value 0.1429. To understand the distribution of estimated doses we have drawn a density plot in figure 11 and observe that the whole distribution of estimated doses are very similar in both methods and the 95% confidence intervals are nearly the same. The only visible difference in both of density plots is that the density at the mean value is slightly higher while using the best fit model. To verify one to one values for each simulation we have drawn the scattered plot in figure 12 and we can see that it resembles the line with slope 1. We can explain this behavior by the reasoning that the posterior weight is higher when the model is close to the observed data and the posterior weight is lower when the model is not that good fit to the observed data so in the weighted average we get the estimate close to the estimate calculated by the best fit model.

Method	Mean value	Standard deviation	95% confidence interval
Estimated using best fit model	0.145	0.025	[0.099, 0.196]
Estimated using weighted average of fitted models	0. 145	0.026	[0.097, 0.196]

Table 6: Estimated dose mean value, standard deviation, and 95% Confidence interval using optimal design

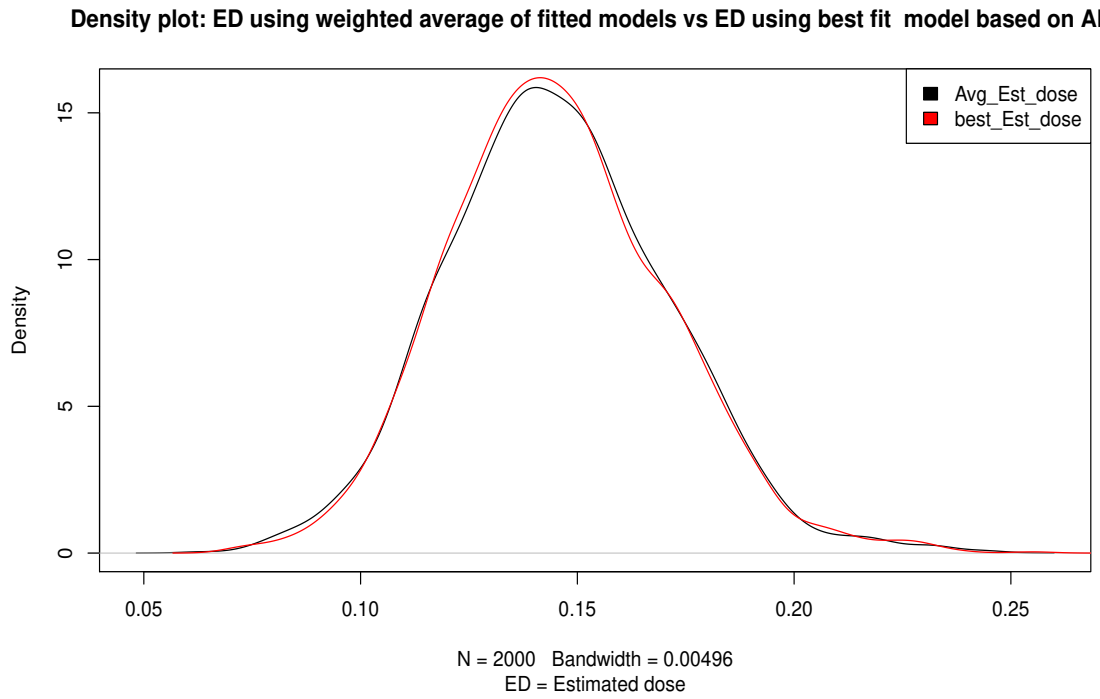

Figure 11: Density plot for estimated dose values using different methods

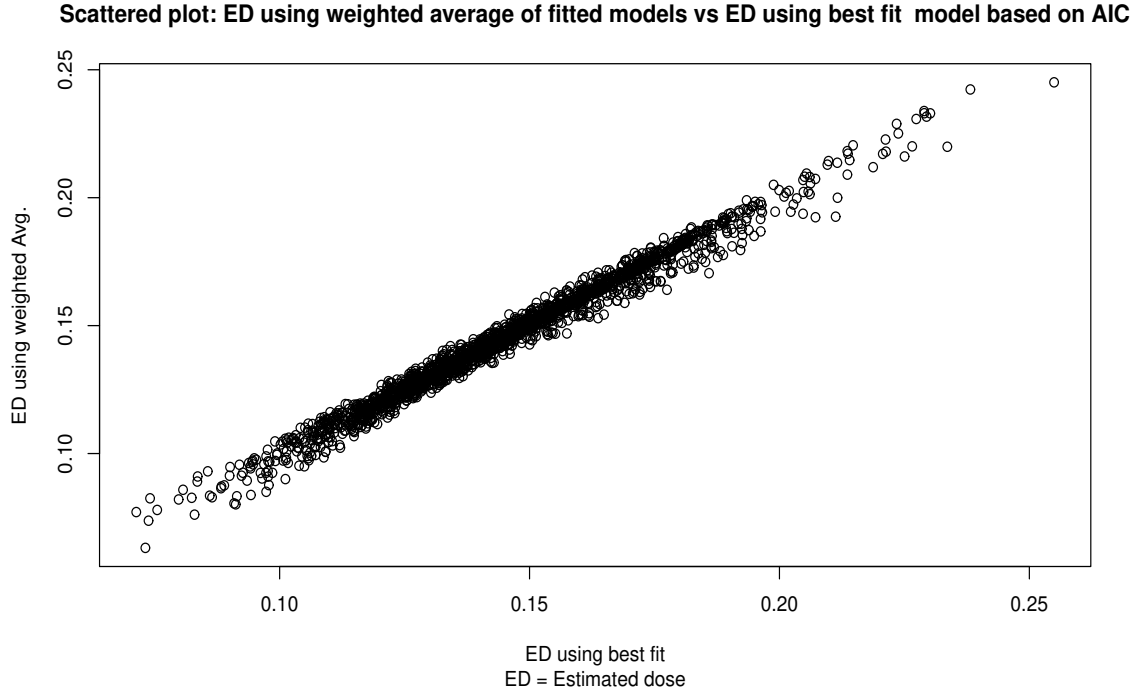

Figure 12: Scattered plot for estimated dose values using best fit model and using weighted average of all fitted models

5.3 Toxicity detection

In the context of pharmacology, drug toxicity occurs when a person has accumulated too much of a drug in his bloodstream, leading to adverse effects on the body. Usually in toxic response, we see the maximum response at dose below to the highest dose and then decrease in the response. As we have discussed in section 2.1 this type of dose-response behavior can be captured by using the betaMod model in which we can define the maximum effect at any dose below to the highest dose. So we can generate data points using the betaMod model which resemble the toxic response of few subjects.

In this experiment our objective is to check that by using the MCPMod methodology we can identify if there are some subjects that have a toxic response. To do that we have used the same experimental design which we have used for the previous experiments that is same set of candidate models with sample size [7,2,3,4,4,7] for doses [0,0.1,0.15,0.35,0.7,1] respectively. As discussed in section 4.2.1 in our candidate models we have already added one betaMod model in the hope that this can help us in detecting toxicity.

To do the analysis at different levels of toxicity, we have generated data sets by merging the two data sets. One generated using the Emax model with parameters $E_0 = 0$, $E_{max} = 1.2$, $ED_{50} = 0.2$ and another generated using the betaMod model with parameters $E_0 = 0$, $E_{max} = 1$, $\delta_1 = 2.54$ and $\delta_1 = 1.52$. It is to be noted that the betaMod model parameters are the same as in our candidate betaMod model. In these two data sets data samples from the Emax model represent the expected response and data samples from the betaMod samples represent the samples with toxic response. By merging the datasets we have simulated the data set with expected responses mixed with toxic responses at higher doses.

Table 7 describe in detail how we have generated different types of data sets. For all these types of merged data sets we have generated data for 2000 simulations and used in the analysis.

Level of toxicity in data	Data set 1	Data set 2
No toxic response samples	[7,2,3,4,4,7]	No samples
1 toxic response sample at two highest doses	[7,2,3,4,3,6]	[0,0,0,0,1,1]
2 toxic response samples at two highest dose	[7,2,3,4,2,5]	[0,0,0,0,2,2]
3 toxic response samples at two highest dose	[7,2,3,4,1,4]	[0,0,0,0,3,3]
4 toxic response samples at two highest dose	[7,2,3,4,0,3]	[0,0,0,0,4,4]
4 toxic response samples at second highest dose and 5 at highest dose	[7,2,3,4,0,2]	[0,0,0,0,4,5]

Table 7: Data set 1: Emax model with parameters $E_0 = 0$, $E_{max} = 1.2$, $ED_{50} = 0.2$ and Data set 2: betaMod model with parameters $E_0 = 0$, $E_{max} = 1$, $\delta_1 = 2.54$ and $\delta_1 = 1.52$ (Sample sizes for doses [0,0.1,0.15,0.35,0.7,1] respectively)

For each level of toxicity in data and for each simulation we have run the analysis using the *MCPMod* function of package *DoseFinding* [14] and calculated the betaMod posterior weight. Depending on the observed data set we have got the value of the betaMod posterior weight ranging from 0 to 1. It is not possible to quantify the expected betaMod posterior weight based on the number of toxic samples. To understand how the betaMod posterior weight varies with the increase in toxic samples we have generated the histogram and density plot for each level of toxicity. This can be seen in the figures 13, 14, 15. We can see that as toxicity increases, the distribution of the betaMod posterior weight shifts towards 1 and we can say that the probability of getting a higher betaMod posterior weight increases. This increase is significant in cases when there are many toxic samples for example as in figure 15.

There is no clear definition for when a dose level is to be considered toxic or not based on the posterior weights. So it is hard to define a value of the betaMod posterior weight at which we say there is toxicity. One possible way of detecting the toxicity could be to set a threshold for a value of the betaMod posterior weight in the design and say that if we get a betaMod posterior weight higher than that value we will conclude that there are toxic responses in a few samples and that we

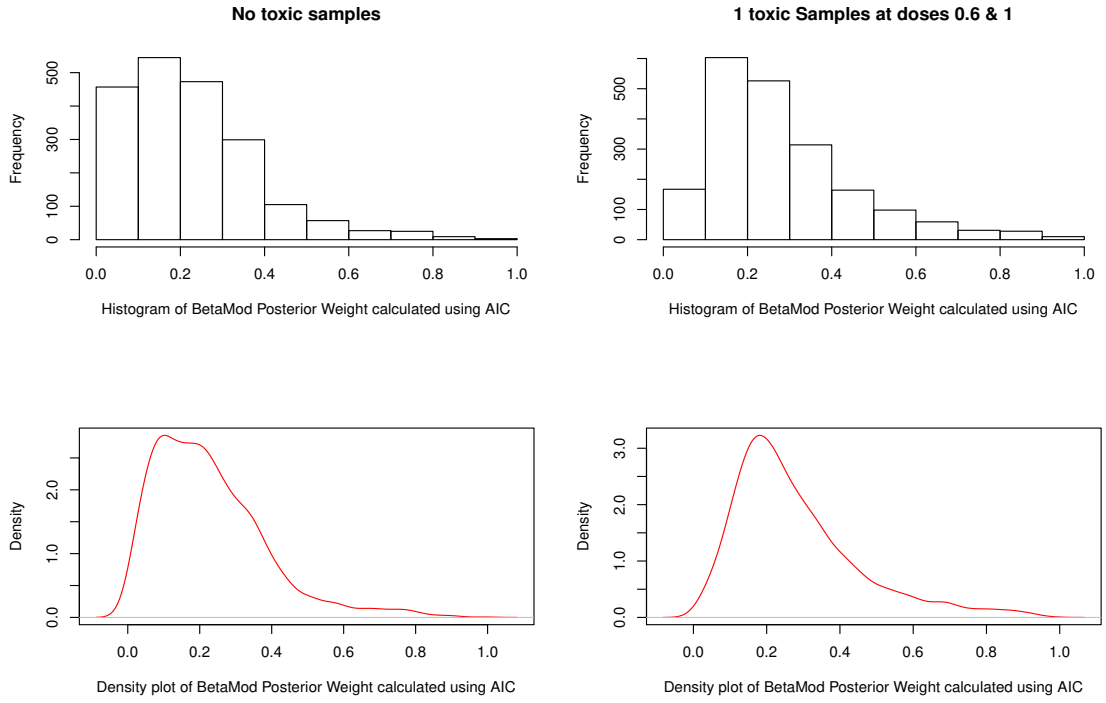

Figure 13: Histogram and density plot for betaMod posterior weight with no or very few toxic response samples

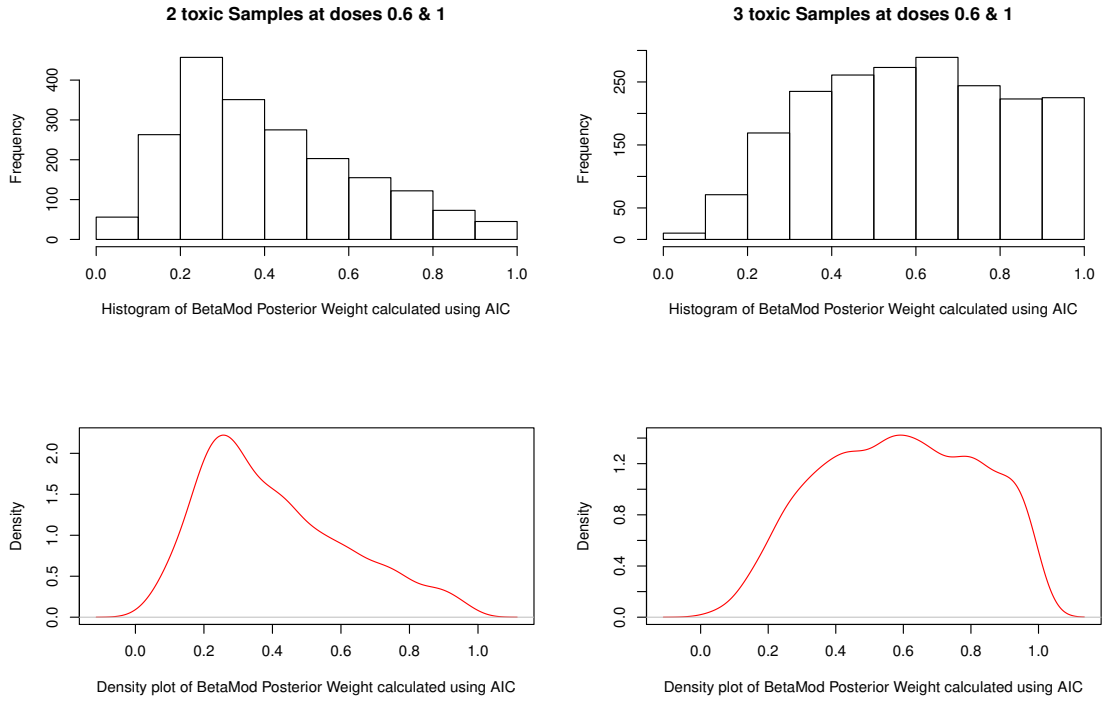

Figure 14: Histogram and density plot for betaMod posterior weight with few toxic response samples

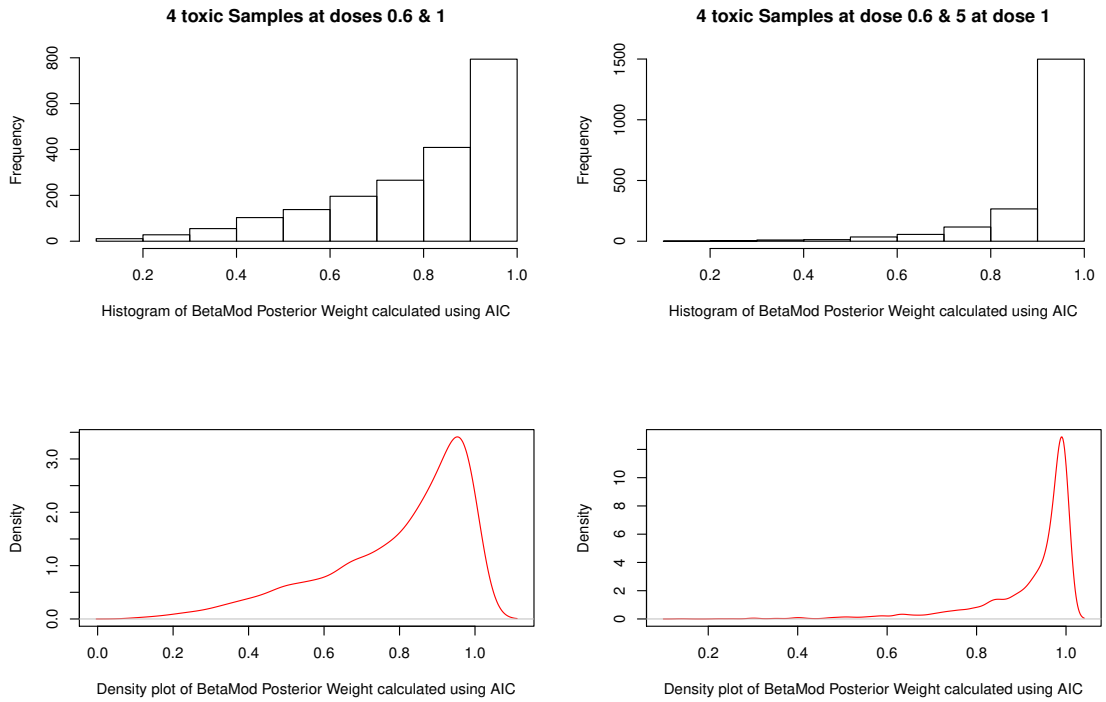

Figure 15: Histogram and density plot for betaMod posterior weight with many toxic response samples

have detected the toxicity. We have set this threshold value to 0.5 and calculated the percentage of experiments we detect the toxicity for each level of toxicity presented in simulated data set. The results are reported in table 8. We can see that as the number of toxic response samples increases the percentage of detected toxicity also increases and in case of very high numbers of toxic response samples we can almost surely detect the toxicity.

Type of data	Count	Detected toxicity Percentage
No toxic response samples	121	6.05
1 toxic response sample at two highest doses	226	11.3
2 toxic response samples at two highest dose	598	29.9
3 toxic response samples at two highest dose	1254	62.7
4 toxic response samples at two highest dose	1803	90.15
4 toxic response samples at second highest dose and 5 at highest dose	1973	98.65

Table 8: Count of the betaMod posterior weight > 0.5 out of 2000 simulations and the percentage of experiments in which toxicity is detected for all toxicity levels

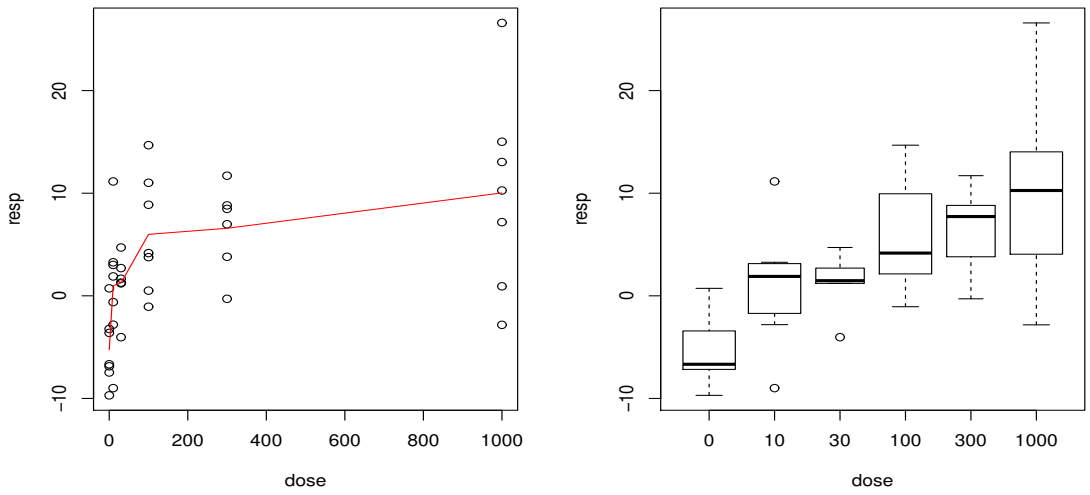

Figure 16: Dose-response data for previous study with drug A (Response: Ejection Fraction and doses [0,10,30,100,300,1000] , red line is joining the mean value of response at each dose level

6 Application: Design based on real data

This section is based on observed data from a previous study in which drug A is used with dose levels [0,10,30,100,300,1000] for the treatment of Myocardial Infarct(MI) in mice and in this study the primary endpoint for response was ejection fraction, reflecting the heart capacity. Full data is presented in section A.3. Based on this data we attempt to design a future experiment with Drug B. In vitro data indicate they have similar potency and our aim of the study is to understand how the in vivo potency of a new compound compares with drug A with a minimal design. We will neglect the PK differences and assume that both drugs have the same PK.

We want to place one dose greater than and one smaller than the ED_{50} for Drug B. The closer the better in terms of the precision of our estimate/interval, but with high variability we will struggle to resolve differences for a narrow interval. Additionally there is a chance that drug B is more or less potent than drug A. Using MCPMod we attempt to find the appropriate dose levels for future studies with Drug B.

Figure 16 shows the given dose- response data and the respective box plot, we can see that the variance in the response is high and varied across all dose levels. It is one of the types of experiments in which we observe high variability in pre-clinical experiments. This is due to the fact that in the response of this experiment we measure heart function in mice after myocardial infarction and there is inherent variability in functional cardiac parameters after induction of heart failure by MI in mice.[35]

First we attempted to fit this data using the MCPMod method. From figure 16 we notice that the dose-response shape is similar to the Emax model so we have used four Emax models as candidate models. To evaluate the parameters for the standardized version of each of these models we have used the values at different dose levels. For each dose level we have calculated percentage effects at each dose by dividing the change in mean response at that dose by the change in mean response at the maximum dose. Table 9 shows the value of percentage effects at each dose and corresponding ED_{50} parameters. To calculate the full candidate models we have used mean response at dose level 0 as control effect and mean response at the maximum dose level as maximum effect.

Model Name	Value from given data	Output parameters
Emax1	per = 0.77 at d = 300	ED_{50} = 89.61
Emax2	per = 0.74 at d = 100	ED_{50} = 35.14
Emax3	per = 0.43 at d = 30	ED_{50} = 39.77
Emax4	per = 0.41 at d = 10	ED_{50} = 14.4

Table 9: Parameters for standardized candidate models and used value from given data(per is fraction of maximum effect and d is dose level)

We have used the above decided candidate models to conduct one-sided hypothesis testing at significance level= 0.05,run the analysis using the *MCPMod* function and got the following results(The full results are presented in section A.4):

```
Multiple Contrast Test:
         t-Stat  adj-p
Multiple Contrast Test:
      t-Stat  adj-p
emax2  5.178 <0.001
emax3  5.170 <0.001
emax4  5.138 <0.001
emax1  5.055 <0.001

Critical value: 1.857 (alpha = 0.05, one-sided)
```

In the first step of the MCPMod method(Multiple comparison procedure), all the selected candidate models were significant and we can see that for all the models the multiplicity adjusted p-values are less than 0.001 and corresponding t-statistic values are significantly higher than the t-critical value of 1.857.

Figure 17 shows the fitted model with 95% confidence interval. Table 10 describes the estimate, standard error, and 95% confidence intervals for all estimated parameters for the fitted model and we can observe that for the fitted Emax model, the estimated value for ED_{50} parameter is 27.25 but there is a very high standard error and in turn this has a very wide 95% confidence interval. This means that this parameter is not identifiable from the given data. This could be due to the large variability in the given data set.

Model name	Estimated Parameters	Standard Error	95% CI for Parameters
Emax	e0 = -4.53	2.11	e0: [-8.66,-0.41]
	eMax = 13.60	2.58	eMax : [8.55,18.65]
	ed50 = 27.25	20.28	ed50: [-12.5,67.00]

Table 10: Parameters for the fitted model with previous experimental data using drug A

Now based on the above findings we attempt to design the experiment for drug B. At first, we have decided the values for the candidate models based on the parameter estimates of the fitted model. Considering the wide confidence interval ED_{50}, we have assumed four candidate models with ED_{50} value as 5, 25,50,75. The estimates of E_0 and E_{max} parameters has smaller standard errors so We have assumed the estimate of the E_0 parameter as control effect and the difference of estimates of E_{max} and E_0 parameters as the maximum effect. These assumptions gave us the candidate models as described in table 11 and we have guessed initial doses as [0,10,20,30,40,50,60,70,80,90,100,200,300,500,700,1000]. We have selected more doses in the range of 0 to 100 as we can see from the fitted model that the main change in the shape happens in this region of the dose levels and our aim is to find doses around possible ED_{50} of drug B.

To find out the optimal doses and the respective allocation ratio we have used the function *optDesign* of the *DoseFinding* [14] package . This function can calculate the optimal design for estimating

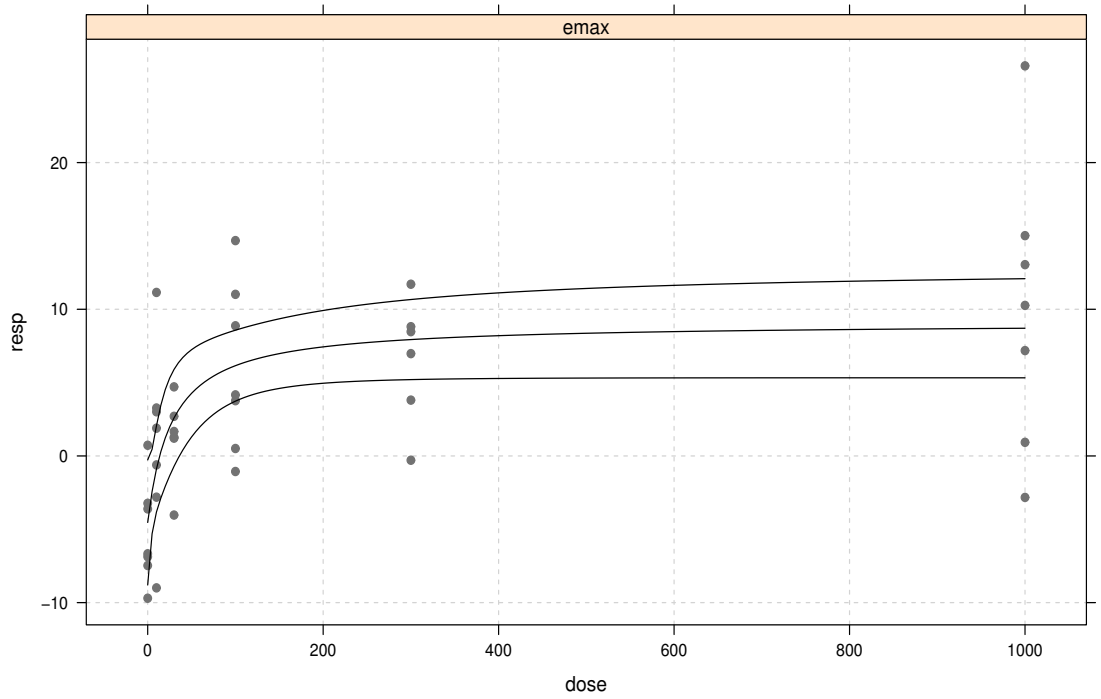

Figure 17: Fitted models along with 95% confidence interval with all data points of given data set of ejection fraction for dose levels [0,10,30,100,300,1000]

Candidate Models	$f(d,\theta)$
Emax1	$-4.53 + \frac{d \cdot 15.37}{d+5}$
Emax2	$-4.53 + \frac{d \cdot 15.67}{d+25}$
Emax3	$-4.53 + \frac{d \cdot 16.05}{d+50}$
Emax4	$-4.53 + \frac{d \cdot 16.44}{d+75}$

Table 11: All candidate models for future experiment with drug B

the dose-response model parameters (D-optimal). We have provided all candidate models given in table 11 with equal probabilities and calculated the dose levels and allocation ratios. Table 12 shows the results and it suggests the dose levels 0, 10 , 50, and 1000. Based on these results we can recommend to use dose level 10 and dose level 50. However from the given data we were not able to identify the ED_{50} parameter due to a wide confidence interval. Considering this, we can not deduce where the selected dose levels [10,50] stand with respect to ED_{50} of drug B. If we consider the estimate of ED_{50} for drug A, which was 27.25 then the selected dose level 10 is below this value and selected dose level 50 is above this value. Additionally, if we are allowed to use one more dose level we would like to recommend dose level 100 as from the fitted model in figure 17, we can observe that at this level shape of fitted model is changed from increasing to flat.

Dose levels	Allocation ratio
0	0.31
10	0.15
50	0.23
1000	0.31

Table 12: Dose levels and respective allocation ratio for future experiment with drug B

7 Discussion and Outlook

The MCP-Mod methodology allows the combination of multiple comparisons and modeling methods in a statistically principled way. Such methodology is useful for Phase II clinical studies, when one is interested in testing the presence of dose-response and in selecting the adequate dose(s) for the confirmatory phase of development.[15] One of the clear advantages of this method in comparison to more traditional approaches, is its added flexibility in searching for and identifying an adequate dose for future drug development. Instead of fitting one dose relationship we have the option of fitting many possible dose relationship shapes. Also when we use multiple comparison methods we often see that ED_{50} is in between the suggested doses and using this method we can estimate the value of ED_{50}.

In this thesis we have attempted to explore if we can use this methodology for pre-clinical experiments and then tried to answer the important questions related to dose-response studies. We have explored the possibility of using the MCPMod method in improving our experimental design, estimating the underlying model from observed data, estimating dose at the given percentage of maximum observed effect, and identifying the toxicity in responses at high doses.

While designing our experiments we have the choice of selecting a normal design when sample sizes are equal for each dose level or an optimal design when sample sizes are not equal for each dose level but decided based on the specified allocation ratio. If results are not negatively affected it is recommended to use optimal design as we can save animals and other resources such as total dose. In our design we saw that we could reduce the sample size by 25% as 27 samples were required in optimal design and 36 samples were required in using normal design. Also we have noted that this method can also be useful in identifying the required dose levels for experiments.

We have simulated data as specified in the design and run the analysis using the MCPMod method and observed that we were able to estimate the underlying model of simulated data and the resulting parameters estimates were more close to the true estimates in case of optimal design. It is to be noted that data is separately simulated for both designs so results were not similar but we observe that in both cases the estimates are close to the true estimates and had similar standard errors. Based on these results we can say that we can use the MCPMod methodology for design and analysis of pre-clinical experiments for dose-response studies. We can say that results are not affected negatively by using optimal design so this is recommended as it allows us to save resources.

After this we have conducted the analysis for 2000 simulations and found that probability of detecting the underlying model as best fitted model is a lot higher as compare to other candidate models. This probability will increase if we have candidate models which are not similar to the underlying model. For all simulations we have calculated the estimated dose at 50% of observed maximum effect using two methods (i) using best fitted model (ii) using weighted average of all fitted models where weights has been calculated based on AIC values. We have observed that estimates using both methods were very similar to each other. This suggests, we can use weighted average of all the fitted models to estimate the dose in place of best fitted model as this method consider all the fitted candidate models and model averaging techniques offer the possibility to take into account the uncertainty in the model selection process and prevent potential bias.

For toxicity analysis, we have simulated data ranging from no samples with toxic response to many samples with toxic response at higher doses. For each level of toxicity we have simulated and analysed 2000 simulations. We have observed that the frequency of observing higher posterior weights of the BetaMod model is increased when toxicity level is increased and when the level of toxicity is very high the posterior weight of the BetaMod model is mostly very high. This suggests that by adding one BetaMod model to the candidate models in the design, we can identify toxicity at higher levels.

Finally we have attempted to apply the methodology for designing future experimental studies for a different drug using the historical data of another drug for the same disease. We have been

given the data of a dose-response study conducted using drug A for treating Myocardial Infarct in Mice. The measured response were ejection fraction, reflecting the heart capacity. In this case we have found that data has high variability and varied responses at different dose levels while in all our experiments we have assumed homoscedastic normality at each dose level. We tried to fit the observed data with a similar methodology and the fitted model has an ED_{50} estimate with a higher standard error and a wider 95% confidence interval. Taking note of this, for future designs we used multiple Emax candidate models with different ED_{50} parameters covering the whole interval. Using these candidate models and providing many dose levels as input we could determine suggested dose levels for future studies with a minimal design.

Overall we can say that we can use the MCPMod methodology in pre-clinical experiments for dose-response studies. This methodology can help us in optimizing the design so that we can reduce the required resources such as animals and dose. This can also reduce the time of study by replacing separate experiments for PD study and efficacy study to single experiment. Additionally this methodology can help us in identifying the important things such as toxicity at earlier stages.

We think there is a lot of potential for future research in this area. In this thesis we have worked on the data in which error was normally distributed with equal variances across all dose levels. We have seen in the example data, in pre-clinical experiments we might observe data with varied variances across dose levels. It would be interesting to explore how can we use MCPMod methodology on this type of data. In this thesis we have used the D-optimal criterion, it would be interesting to see the results after applying other optimally criteria described in section 4.1.1. We think that there is more scope of research in formulating the studies related to toxicity identification such that evaluating the effect of defining the prior weight of the BetaMod model while designing the study. Apart from the MCP-Mod approach, there is a similar Bayesian model strategy (BMA-Mod) for carrying out analyses. Studies of the performance of BMA methods have demonstrated a number of desirable attributes. [36] There is scope to evaluate this method for design and analysis of pre-clinical experiments for dose-response studies.

References

1. Cook, D. *et al.* Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nature reviews Drug discovery* **13,** 419–431 (2014).
2. Bretz, F., Dette, H. & Pinheiro, J. C. Practical considerations for optimal designs in clinical dose finding studies. *Statistics in medicine* **29,** 731–742 (2010).
3. Bretz, F., Pinheiro, J. C. & Branson, M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **61,** 738–748 (2005).
4. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery* **3,** 711–716 (2004).
5. Atkinson Jr, A. & Lalonde, R. Introduction of quantitative methods in pharmacology and clinical pharmacology: a historical overview. *Clinical Pharmacology & Therapeutics* **82,** 3–6 (2007).
6. Mould, D. & Upton, R. N. Basic concepts in population modeling, simulation, and model-based drug development—part 2: introduction to pharmacokinetic modeling methods. *CPT: pharmacometrics & systems pharmacology* **2,** 1–14 (2013).
7. Huang, S.-M., Abernethy, D. R., Wang, Y., Zhao, P. & Zineh, I. The utility of modeling and simulation in drug development and regulatory review. *Journal of pharmaceutical sciences* **102,** 2912–2923 (2013).
8. Upton, R. & Mould, D. Basic concepts in population modeling, simulation, and model-based drug development: part 3—introduction to pharmacodynamic modeling methods. *CPT: pharmacometrics & systems pharmacology* **3,** 1–16 (2014).
9. Mould, D. R. & Upton, R. Basic concepts in population modeling, simulation, and model-based drug development. *CPT: pharmacometrics & systems pharmacology* **1,** 1–14 (2012).
10. Mehrotra, N, Gupta, M, Kovar, A & Meibohm, B. The role of pharmacokinetics and pharmacodynamics in phosphodiesterase-5 inhibitor therapy. *International journal of impotence research* **19,** 253–264 (2007).
11. Derendorf, H. & Meibohm, B. Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: concepts and perspectives. *Pharmaceutical research* **16,** 176–185 (1999).
12. Macdougall, J. in *Dose finding in drug development* 127–145 (Springer, 2006).
13. Team, R. C. *R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2014, Vienna, Austria* 2014.
14. Bornkamp, B., Pinheiro, J., Bretz, F. & Bornkamp, M. B. Package 'DoseFinding' (2019).
15. Pinheiro, J., Bornkamp, B. & Bretz, F. Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. *Journal of biopharmaceutical statistics* **16,** 639–656 (2006).
16. Bornkamp, B., Pinheiro, J., Bretz, F. & Bornkamp, M. B. Package 'MCPMod' (2016).
17. Hsu, J. *Multiple comparisons: theory and methods* (CRC Press, 1996).
18. Ruberg, S. J. Dose response studies I. Some design considerations. *Journal of Biopharmaceutical Statistics* **5,** 1–14 (1995).
19. Henning, K. S. & Westfall, P. H. Closed testing in pharmaceutical research: Historical and recent developments. *Statistics in biopharmaceutical research* **7,** 126–147 (2015).
20. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics,* 65–70 (1979).
21. Hommel, G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75,** 383–386 (1988).
22. Hochberg, Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75,** 800–802 (1988).
23. Tukey, J. W. The problem of multiple comparisons. *Multiple comparisons* (1953).
24. Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50,** 1096–1121 (1955).
25. Krzykalla, J. *Investigations on MCP-Mod Designs* PhD thesis (2015).
26. Bates, D. M. & Watts, D. G. *Nonlinear regression analysis and its applications* (Wiley New York, 1988).
27. Bates, D., Chambers, J. & Hastie, T. *Statistical models in S* in *Computer science and statistics: proceedings of the 19th Symposium on the Interface. Wadsworth & Brooks, California* (1992).

28. Pinheiro, J. C. & Bates, D. M. Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus,* 3–56 (2000).
29. Pinheiro, J. C., Bretz, F. & Branson, M. in *Dose finding in drug development* 146–171 (Springer, 2006).
30. Tukey, J., Ciminera, J. & Heyse, J. Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics,* 295–301 (1985).
31. Genz, A. & Bretz, F. Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics* **11,** 950–971 (2002).
32. Bornkamp, B. *Comparison of model-based and model-free approaches for the analysis of dose-response studies* PhD thesis (Master's thesis, Fachbereich Statistik, Universität Dortmund. URL http://www . . ., 2006).
33. Buckland, S. T., Burnham, K. P. & Augustin, N. H. Model selection: an integral part of inference. *Biometrics,* 603–618 (1997).
34. Russell, W. M. S. & Burch, R. L. *The principles of humane experimental technique* (Methuen, 1959).
35. Chen, J. *et al.* The probability of inconstancy in assessment of cardiac function post-myocardial infarction in mice. *Cardiovascular pharmacology: open access* **5** (2016).
36. Gould, A. L. BMA-Mod: A Bayesian model averaging strategy for determining dose-response relationships in the presence of model uncertainty. *Biometrical Journal* **61,** 1141–1159 (2019).

A Supplementary material

A.1 Simulated data for experiments

Full data Generated for dose levels [0,0.1,0.15,0.35,0.7,1] using E_{max} model with parameter E_0 = 0, E_{max} = 1.2, ED_{50} = 0.2, with following set up:
Sample sizes [7,2,3,4,4,7] and between subject standard deviation = 0.1 at each dose level

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
dose	0.00000000	0.1000000	0.1500000	0.3500000	0.7000000	1.0000000
1	0.05855288	0.3723816	0.4223535	0.8006992	0.8446976	1.0779622
2	0.07094660	0.3715840	0.5026609	0.8156580	0.9001756	1.1455785
3	-0.01093033		0.6960169	0.6885832	1.0454046	0.9355672
4	-0.04534972			0.8453263	0.9632057	0.8446863
5	0.06058875					0.8402290
6	-0.18179560					1.1805098
7	0.06300986					0.9518353

Full data Generated for dose levels [0,0.1,0.15,0.35,0.7,1] using E_{max} model with parameter E_0 = 0, E_{max} = 1.2, ED_{50} = 0.2, with following set up:
Sample sizes [6,6,6,6,6,6] and between subject standard deviation = 0.1 at each dose level

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
dose	0.00000000	0.1000000	0.1500000	0.3500000	0.7000000	1.000000
1	0.05855288	0.4630099	0.5513485	0.8757076	0.7735624	1.081187
2	0.07094660	0.3723816	0.5663074	0.7935087	1.1138431	1.219683
3	-0.01093033	0.3715840	0.4392325	0.8415986	0.8851686	1.204919
4	-0.04534972	0.3080678	0.5959757	0.9092149	0.9953713	1.163245
5	0.06058875	0.3883752	0.4256500	0.6992035	0.9945457	1.025427
6	-0.18179560	0.5817312	0.4811280	0.6083226	0.9171022	1.049119

A.2 Full results of experiments in section 5.1

Below is full output summary of results after analysing simulated data for normal design with sample sizes [7,2,3,4,4,7] at dose levels [0,0.1,0.15,0.35,0.7,1] using MCPMod method.

```
MCPMod

**************************************
MCP part
**************************************
Multiple Contrast Test

Contrasts:
      emax1  emax2  emax3 betaMod sigEmax logistic
0    -0.824 -0.787 -0.752  -0.628  -0.664   -0.556
0.1  -0.048 -0.078 -0.097  -0.168  -0.137   -0.142
0.15 -0.008 -0.053 -0.085  -0.224  -0.156   -0.196
0.35  0.150  0.122  0.092   0.013   0.022   -0.111
0.7   0.244  0.256  0.260   0.604   0.268    0.264
1     0.487  0.540  0.584   0.403   0.667    0.742

Contrast Correlation:
         emax1 emax2 emax3 betaMod sigEmax logistic
emax1    1.000 0.994 0.980   0.821   0.925    0.835
emax2    0.994 1.000 0.996   0.857   0.962    0.889
emax3    0.980 0.996 1.000   0.876   0.982    0.925
betaMod  0.821 0.857 0.876   1.000   0.894    0.867
```

```
sigEmax  0.925 0.962 0.982   0.894   1.000    0.977
logistic 0.835 0.889 0.925   0.867   0.977    1.000

Multiple Contrast Test:
         t-Stat  adj-p
emax2    19.289 <0.001
emax1    19.223 <0.001
emax3    19.164 <0.001
sigEmax  18.393 <0.001
logistic 16.850 <0.001
betaMod  16.426 <0.001

Critical value: 2.031 (alpha = 0.05, one-sided)

****************************************
Mod part
****************************************
** Fitted model 1
Dose Response Model

Model: emax
Fit-type: normal

Residuals:
    Min      1Q  Median      3Q     Max
-0.1792 -0.0672  0.0107  0.0576  0.1816

Coefficients with approx. stand. error:
      Estimate Std. Error
e0   -0.000207     0.0382
eMax  1.190175     0.0748
ed50  0.188344     0.0440

Residual standard error: 0.102
Degrees of freedom: 24

** Fitted model 2
Dose Response Model

Model: betaMod
Fit-type: normal

Residuals:
    Min      1Q  Median      3Q     Max
-0.1891 -0.0666  0.0113  0.0637  0.1822

Coefficients with approx. stand. error:
       Estimate Std. Error
e0     0.000366     0.0401
eMax   1.005206     0.0517
delta1 0.478728     0.1060
delta2 0.159357     0.0969

Residual standard error: 0.106
Degrees of freedom: 23
```

```
** Fitted model 3
Dose Response Model

Model: sigEmax
Fit-type: normal

Residuals:
    Min      1Q  Median      3Q     Max
-0.1870 -0.0601  0.0127  0.0525  0.1835

Coefficients with approx. stand. error:
     Estimate Std. Error
e0    0.00173     0.0390
eMax  1.09683     0.1441
ed50  0.16245     0.0419
h     1.23522     0.4515

Residual standard error: 0.103
Degrees of freedom: 23

** Fitted model 4
Dose Response Model

Model: logistic
Fit-type: normal

Residuals:
    Min      1Q  Median      3Q     Max
-0.2093 -0.0500 -0.0017  0.0593  0.1936

Coefficients with approx. stand. error:
      Estimate Std. Error
e0      -0.942     1.5521
eMax     1.914     1.5673
ed50     0.001     0.2153
delta    0.135     0.0619

Residual standard error: 0.108
Degrees of freedom: 23

****************************************
Model selection criteria (aveAIC):
****************************************
     emax   betaMod   sigEmax  logistic
-42.01810 -38.81453 -40.33912 -37.95860

Model weights (AIC):
    emax  betaMod  sigEmax logistic
  0.5666   0.1142   0.2447   0.0744

****************************************
Estimated ED, p=0.5
****************************************
    emax  betaMod  sigEmax logistic
  0.1368   0.1389   0.1390   0.1489
```

Below is full output summary of results after analysing simulated data for normal design with sample sizes [6,6,6,6,6,6] at dose levels [0,0.1,0.15,0.35,0.7,1] using MCPMod method.

```
 MCPMod

**************************************
MCP part
**************************************
Multiple Contrast Test

Contrasts:
      emax1  emax2  emax3 betaMod sigEmax logistic
0    -0.780 -0.707 -0.648  -0.387  -0.505   -0.383
0.1  -0.180 -0.248 -0.286  -0.361  -0.352   -0.335
0.15 -0.043 -0.113 -0.161  -0.318  -0.257   -0.302
0.35  0.215  0.188  0.155   0.043   0.079   -0.090
0.7   0.366  0.398  0.413   0.728   0.435    0.442
1     0.421  0.482  0.527   0.295   0.600    0.668

Contrast Correlation:
         emax1 emax2 emax3 betaMod sigEmax logistic
emax1    1.000 0.990 0.970   0.781   0.897    0.795
emax2    0.990 1.000 0.995   0.839   0.950    0.868
emax3    0.970 0.995 1.000   0.869   0.977    0.913
betaMod  0.781 0.839 0.869   1.000   0.901    0.880
sigEmax  0.897 0.950 0.977   0.901   1.000    0.975
logistic 0.795 0.868 0.913   0.880   0.975    1.000

Multiple Contrast Test:
         t-Stat  adj-p
emax2    22.905 <0.001
emax3    22.859 <0.001
emax1    22.592 <0.001
sigEmax  21.991 <0.001
logistic 20.308 <0.001
betaMod  18.744 <0.001

Critical value: 2.018 (alpha = 0.05, one-sided)

**************************************
Mod part
**************************************
** Fitted model 1
Dose Response Model

Model: emax
Fit-type: normal

Residuals:
     Min       1Q   Median       3Q      Max
-0.19018 -0.07122  0.00216  0.05978  0.22355

Coefficients with approx. stand. error:
     Estimate Std. Error
e0   -0.00019     0.0391
eMax  1.34343     0.0715
```

```
ed50  0.24294     0.0437

Residual standard error: 0.0978
Degrees of freedom: 33

** Fitted model 2
Dose Response Model

Model: betaMod
Fit-type: normal

Residuals:
     Min       1Q   Median       3Q      Max
-0.16453 -0.07287 -0.00218  0.06173  0.21538

Coefficients with approx. stand. error:
       Estimate Std. Error
e0     -0.00969     0.0401
eMax    1.12677     0.0801
delta1  0.44294     0.0720
delta2  0.05000     0.0717

Residual standard error: 0.0985
Degrees of freedom: 32

** Fitted model 3
Dose Response Model

Model: sigEmax
Fit-type: normal

Residuals:
     Min       1Q   Median       3Q      Max
-0.16389 -0.07187 -0.00312  0.06281  0.21251

Coefficients with approx. stand. error:
     Estimate Std. Error
e0   -0.00869     0.0396
eMax  1.99099     1.1326
ed50  0.70154     1.2093
h     0.66712     0.2574

Residual standard error: 0.0971
Degrees of freedom: 32

** Fitted model 4
Dose Response Model

Model: logistic
Fit-type: normal

Residuals:
    Min      1Q  Median      3Q     Max
-0.2335 -0.0627 -0.0161  0.0505  0.2444

Coefficients with approx. stand. error:
```

```
      Estimate Std. Error
e0      -0.957     1.2926
eMax     1.997     1.3139
ed50     0.001     0.1931
delta    0.156     0.0591

Residual standard error: 0.115
Degrees of freedom: 32

****************************************
Model selection criteria (aveAIC):
****************************************
     emax   betaMod   sigEmax  logistic
-60.38221 -58.97291 -59.95446 -47.63632

Model weights (AIC):
    emax  betaMod  sigEmax logistic
  0.4341   0.2146   0.3505   0.0007

****************************************
Estimated ED, p=0.5
****************************************
    emax  betaMod  sigEmax logistic
  0.1635   0.1773   0.1696   0.1710
```

A.3 Historical data of previous study

Full dose-response data of previous study in which drug A is used with dose levels [0,10,30,100,300,1000] for the treatment of Myocardial Infarct(MI) in mice and in this study primary endpoint for response was ejection fraction, reflecting capacity of heart.

```
       [,1]      [,2]      [,3]       [,4]       [,5]        [,6]
dose  0.000 10.000000 30.000000 100.000000 300.000000 1000.000000
1     0.728 -8.994596  2.702602   8.874409   6.975372    7.183406
2    -3.230 -2.809798  4.709994   0.506821  -0.290615   13.042204
3    -6.670  1.891771  1.212081  14.681336   8.474189   10.264326
4    -3.610  3.011568  1.286021   4.164526   8.812609   26.594684
5    -9.700  3.268864 -4.029345  -1.060348  11.708344   -2.827373
6    -6.870 -0.612570  1.666271   3.767455   3.806658    0.928837
7    -7.470 11.147938             11.018046              15.018660
```

A.4 Full results after fitting historical data

Below is full output summary of results after analysing historical data:

```
 MCPMod

****************************************
MCP part
****************************************
Multiple Contrast Test

Contrasts:
      emax1  emax2  emax3  emax4
0    -0.526 -0.641 -0.624 -0.761
10   -0.401 -0.378 -0.386 -0.275
```

```
30   -0.184 -0.081 -0.097  0.035
100   0.129  0.236  0.226  0.276
300   0.369  0.360  0.364  0.318
1000  0.613  0.504  0.518  0.408

Contrast Correlation:
      emax1 emax2 emax3 emax4
emax1 1.000 0.976 0.982 0.905
emax2 0.976 1.000 1.000 0.974
emax3 0.982 1.000 1.000 0.966
emax4 0.905 0.974 0.966 1.000

Multiple Contrast Test:
      t-Stat  adj-p
emax2  5.178 <0.001
emax3  5.170 <0.001
emax4  5.138 <0.001
emax1  5.055 <0.001

Critical value: 1.851 (alpha = 0.05, one-sided)

****************************************
Mod part
****************************************
Dose Response Model

Model: emax
Fit-type: normal

Residuals:
    Min      1Q  Median      3Q     Max
-17.889  -3.024   0.409   2.524  11.533

Coefficients with approx. stand. error:
     Estimate Std. Error
e0      -4.53       2.11
eMax    13.60       2.58
ed50    27.25      20.28

Residual standard error: 5.83
Degrees of freedom: 37

****************************************
Model selection criteria (AIC):
****************************************
    emax
259.4512

Selected model: emax

****************************************
Estimated ED, p=0.5
****************************************
   emax
25.8381
```