# MCPMod: An **R** Package for the Design and Analysis of Dose-Finding Studies

**Björn Bornkamp**
Technische Universität
Dortmund

**José Pinheiro**
Novartis Pharmaceuticals

**Frank Bretz**
Novartis Pharma AG

### Abstract

In this article the **MCPMod** package for the R programming environment will be introduced. It implements a recently developed methodology for the design and analysis of dose-response studies that combines aspects of multiple comparison procedures and modeling approaches (Bretz *et al.* 2005). The **MCPMod** package provides tools for the analysis of dose finding trials, as well as a variety of tools necessary to plan an experiment to be analyzed using the MCP-Mod methodology.

*Keywords*: clinical trial, dose-response, minimum effective dose, multiple contrast test, phase II trials.

## 1. Introduction

In pharmaceutical drug development, dose-response studies typically have two main goals. The first goal is to establish that changes in dose lead to desirable changes in the (efficacy and/or safety) endpoint(s) of interest, the so-called *proof-of-concept* (PoC) step. Once such a dose-response signal has been shown, the second goal is then to select one or more "good" dose level(s) for the confirmatory Phase III studies, the so-called *dose-finding* step.

Traditionally these goals have been addressed *either* by using a multiple comparison procedure (MCP), *or* by using a modeling (Mod) approach. The MCP approach regards the dose as a qualitative factor and generally makes few, if any, assumptions about the underlying dose-response relationship. However, inferences about the target dose are restricted to the discrete, possibly small, set of doses used in the trial. Within the modeling approach, a parametric (typically non-linear) functional relationship is assumed between dose and response. The dose is taken to be a quantitative factor, allowing greater flexibility for target dose estimation. The validity of the modeling approach, however, strongly depends on an appropriate dose-response
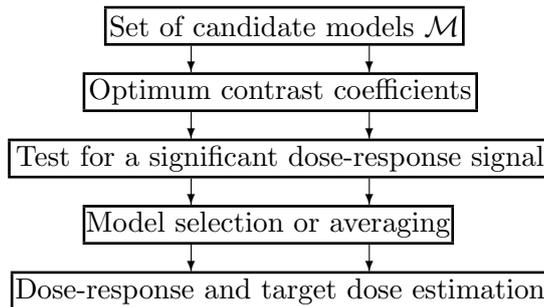
Figure 1: Schematic overview of the MCP-Mod procedure.

model being pre-specified for the analysis.

In this paper we present the **MCPMod** package written in the R system for statistical computing (R Development Core Team 2008) and available from the Comprehensive R Archive Network at `http://CRAN.R-project.org/package=MCPMod`. The package implements a hybrid methodology, combining multiple comparison procedures with modeling techniques (called MCP-Mod procedure, Bretz *et al.* 2005). This approach provides the flexibility of modeling for dose estimation, while preserving the robustness to model misspecification associated with MCP. Figure 1 gives an overview of the MCP-Mod procedure. It starts by defining a set of *candidate* models $\mathcal{M}$ covering a suitable range of dose-response shapes. Each of the dose-response shapes in the candidate set is tested using appropriate contrasts and employing MCP techniques that preserve the family-wise error rate (FWER). PoC is established when at least one of the model tests is significant. Once PoC is verified, either a "best" model or a weighted average of the set of significant models $\mathcal{M}^* \subseteq \mathcal{M}$ is used to estimate the dose-response profile and the target doses of interest.

As outlined above, the MCP-Mod procedure is performed in several steps: (1) calculation of contrast coefficients, representing the candidate model shapes, (2) conduct of a multiple contrast test, and, depending on the result, (3) a model selection step to fit (typically non-linear) dose-response models and to estimate the target doses. Each individual step above can be implemented with the R statistical language, possibly using add-on packages available at `http://CRAN.R-project.org/`. However, it is desirable to have one package, which performs these steps automatically and also allows to design a trial for the MCP-Mod procedure. The **MCPMod** package provides these functionalities and the aim of this paper is to give a detailed description of the package.

For self containment of the paper we will first review the key features and statistical methods of the MCP-Mod procedure in Section 2, while the **MCPMod** package will be introduced and illustrated with examples in Section 3.

## 2. MCP-Mod: Combining multiple comparisons and modeling

### 2.1. Notation

Assume that we observe a response $Y$ for a given set of parallel groups of patients corresponding to doses $d_2, d_3, \ldots, d_k$ plus placebo $d_1$, for a total of $k$ arms. For the purpose of testing

PoC and estimating target doses, we consider the one-way layout

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

where $\mu_{d_i} = f(d_i, \boldsymbol{\theta})$ denotes the mean response at dose $d_i$ for some dose-response model $f(d, \boldsymbol{\theta})$, $n_i$ denotes the number of patients allocated to dose $d_i$, $N = \sum_{i=1}^{k} n_i$ is the total sample size, and $\epsilon_{ij}$ denotes the error term for patient $j$ within dose group $i$. Following Bretz *et al.* (2005), we note that most parametric dose-response models $f(d, \boldsymbol{\theta})$ used in practice can be written as

$$f(d, \boldsymbol{\theta}) = \theta_0 + \theta_1 f^0(d, \boldsymbol{\theta}^*), \tag{2}$$

where $f^0(d, \boldsymbol{\theta}^*)$ denotes the standardized model function, parameterized by the vector $\boldsymbol{\theta}^*$. In this parameterization, $\theta_0$ is a location and $\theta_1$ a scale parameter such that only the parameter-vector $\boldsymbol{\theta}^*$ determines the shape of the model function. As seen later, it is sufficient to consider the standardized model $f^0$ instead of the full model $f$ for the derivation of the optimal model contrasts.

## 2.2. MCP-Mod methodology

In this section we review the core elements of the MCP-Mod methodology. We start by considering the basic MCP-Mod procedure for the analysis of a dose-response trial and then focus on design issues. For more information on the basic methodology see Bretz *et al.* (2005), for recommendations regarding the practical implementation and design aspects see Pinheiro *et al.* (2006a).

*Analysis considerations*

The motivation for MCP-Mod is based on the work by Tukey *et al.* (1985), who recognized that the power of standard dose-response trend tests depends on the (unknown) dose-response relationship. They proposed to simultaneously use several trend tests and subsequently to adjust the resulting $p-$values for multiplicity. Bretz *et al.* (2005) formalized this approach and extended it in several ways.

Assume that a set $\mathcal{M}$ of $M$ parameterized candidate models is given, with corresponding model functions $f_m(d, \boldsymbol{\theta}_m), m = 1, \ldots, M$, and parameters $\boldsymbol{\theta}_m^*$ of the standardized models $f_m^0$ (determining the model shapes). For each of the dose-response models in the candidate set we would like to test the hypothesis $H_0^m$: $\mathbf{c}_m^\top \boldsymbol{\mu} = 0$, where $\mathbf{c}_m = (c_{m1}, \ldots, c_{mk})^\top$ is the optimal contrast vector representing model $m$, subject to $\sum_{i=1}^{k} c_{mi} = 0$. Each of the dose-response models in the candidate set is hence tested using a single contrast test,

$$T_m = \frac{\sum_{i=1}^{k} c_{mi} \bar{Y}_i}{S \sqrt{\sum_{i=1}^{k} c_{mi}^2 / n_i}}, \quad m = 1, \ldots, M,$$

where $S^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / (N - k)$ is the pooled variance estimate. Every single contrast test thus translates into a decision procedure to determine whether the given dose-response shape is statistically significant, based on the observed data.

The contrast coefficients $c_{m1}, \ldots, c_{mk}$ for the $m$-th model are chosen such that they maximize the power to detect the underlying model. It can be shown that these optimal contrast coefficients do not depend on the full parameter vector $\boldsymbol{\theta}_m$ of the model, but only the parameters in its standardized model function $\boldsymbol{\theta}_m^*$, which determine the model shape (see Bretz *et al.* 2005) and the group sample sizes. Letting $(\mu_{m1}^0, \ldots, \mu_{mk}^0)^\top = (f_m^0(d_1, \boldsymbol{\theta}_m^*), \ldots, f_m^0(d_k, \boldsymbol{\theta}_m^*))^\top$, the $i$th entry of the optimal contrast $\mathbf{c}_m$ for detecting the shape $m$ is proportional to

$$n_i(\mu_{mi}^0 - \bar{\mu}), \; i = 1, \ldots, k, \tag{3}$$

where $\bar{\mu} = N^{-1} \sum_{i=1}^k \mu_{mi}^0 n_i$ (Bornkamp 2006, p. 88, Casella and Berger 1990, p. 519). A unique representation of the optimal contrast can be obtained by imposing the regularity condition $\sum_{i=1}^k c_{mi}^2 = 1$.

The final detection of a significant dose-response signal (i.e., demonstrating PoC), is based on the maximum contrast test statistic

$$T_{\max} = \max\{T_1, \ldots, T_M\}.$$

Under the null hypothesis of no dose-response effect $\mu_{d_1} = \ldots = \mu_{d_k}$ and under the distributional assumptions stated in Equation 1, $T_1, \ldots, T_M$ jointly follow a central multivariate $t$ distribution with $N - k$ degrees of freedom and correlation matrix $\boldsymbol{R} = (\rho_{ij})$, where

$$\rho_{ij} = \frac{\sum_{l=1}^k c_{il} c_{jl}/n_l}{\sqrt{\sum_{l=1}^k c_{il}^2/n_l \sum_{l=1}^k c_{jl}^2/n_l}}. \tag{4}$$

Multiplicity adjusted critical values and $p$−values can be calculated using the identity of the sets $[T_{\max} \leq q] = [T_1 \leq q, \ldots, T_M \leq q]$, where $q$ is a real number. As the joint distribution of $(T_1, \ldots, T_M)^\top$ is multivariate $t$, numerical integration routines for evaluation of multivariate $t$ integrals, such as the randomized quasi-Monte Carlo methods of Genz and Bretz (2002) implemented in the R package **mvtnorm** (Genz *et al.* 2009), can be used to compute the desired equicoordinate quantiles of the multivariate $t$ distribution. PoC is hence established if $T_{\max} \geq q_{1-\alpha}$, where $q_{1-\alpha}$ is the multiplicity adjusted critical value at level $1 - \alpha$ (*i.e.*, the equicoordinate $1 - \alpha$ quantile of the corresponding central multivariate t distribution). Furthermore, all dose-response shapes with contrast test statistics larger than $q_{1-\alpha}$ can be declared statistically significant at level $1 - \alpha$ under strong control of the FWER. These models then form a reference set $\mathcal{M}^* = \{M_1, \ldots, M_L\} \subseteq \mathcal{M}$ of $L$ significant models. If no candidate model is statistically significant, the procedure stops indicating that a dose-response relationship can not be established from the observed data (i.e., no PoC).

If PoC has been established, the next step is to estimate the dose-response curve and the target doses of interest. This can be achieved either by selecting a single model out of $\mathcal{M}^*$ or by applying model averaging techniques to $\mathcal{M}^*$. There are different possibilities to select a single dose-response model out of $\mathcal{M}^*$ for target dose estimation. One can base the choice, for example, on the contrast test statistics, i.e., selecting the model corresponding to the maximum contrast test statistic. Standard information criteria like the *AIC* or *BIC* might also be used. The estimate of the model function is then obtained by calculating the least squares estimates for $\boldsymbol{\theta}$. For non-linear models iterative optimization techniques need to be used, such as those implemented in the `nls` function in R. As the non-linear models described

here are partially linear (see Equation 2), this can be exploited in the `nls` function by using the Golub-Pereyra algorithm (see Golub and Pereyra 2003, for a review of these methods). So we only need to derive starting values for the standardized model parameters $\boldsymbol{\theta^*}$. Although we use automatic methods for finding good data-based starting values for $\boldsymbol{\theta^*}$, convergence problems can occur, especially when the number of dose levels used in the trial is small compared to the parameters in the model function. In the case of non-convergence the 'best' of the remaining significant, converging models can be used for dose estimation, if any. An approach to partially overcome these convergence issues is to use box constraints on $\boldsymbol{\theta^*}$. This will be implemented as an alternative in future versions of the package.

Once a dose-response model has been selected, one can proceed to estimate the target dose(s) of interest. One possible choice is the minimum effective dose ($MED$), which is defined as the smallest dose ensuring a clinically relevant and statistically significant improvement over placebo (Ruberg 1995). Formally,

$$MED = \min\{d \in (d_1, d_k] : f(d) > f(d_1) + \Delta\},$$

where $\Delta$ is the clinical relevance threshold. A common estimate for the $MED$ is

$$\widehat{MED} = \min\{d \in (d_1, d_k] : \hat{f}(d) > \hat{f}(d_1) + \Delta, L(d) > \hat{f}(d_1)\}$$

where $\hat{f}(d)$ is the predicted mean response at dose $d$, and $L(d)$ is the corresponding lower bound of the pointwise confidence intervals of level $1 - 2\gamma$. Note that $\widehat{MED}$ corresponds to the $\widehat{MED}_2$ estimator in Bretz *et al.* (2005), who found this estimator to be least biased compared to two other alternative estimates in a simulation study. A different target dose is the $EDp$ which is defined as the smallest dose that gives a certain percentage $p$ of the maximum effect $\delta_{\max}$ observed in $(d_1, d_k]$. Formally,

$$EDp = \min\{d \in (d_1, d_k] : f(d) > f(d_1) + p\delta_{\max}\}, \tag{5}$$

where $\delta_{\max} = f_{max} - f(d_1)$, and $f_{\max} = \max\limits_{d \in (d_1, d_k]} f(d)$. An estimate $\widehat{EDp}$ is obtained by plugging the empirical estimates into the definition (5).

An alternative to selecting a single dose-response model is to apply model averaging techniques and produce weighted estimates across all models in $\mathcal{M}^*$ for a given quantity $\boldsymbol{\psi}$ of interest. In the context of dose-response analysis, the parameter $\boldsymbol{\psi}$ could for example be a target dose ($MED, EDp, \ldots$) or the mean responses at a specific dose $d \in [d_1, d_k]$. Buckland *et al.* (1997) proposed to use the weighted estimate

$$\widehat{\boldsymbol{\psi}} = \sum_\ell w_\ell \widehat{\boldsymbol{\psi}}_\ell,$$

where $\widehat{\boldsymbol{\psi}}_\ell$ is the estimate of $\boldsymbol{\psi}$ under model $\ell$ for given weights $w_\ell$. The idea is thus to use estimates for the final data analysis which rely on the averaged estimates across all $L$ models. Buckland *et al.* (1997) proposed the use of the weights

$$w_\ell = \frac{p_\ell e^{-\frac{IC_\ell}{2}}}{\sum_{j=1}^L p_\ell e^{-\frac{IC_j}{2}}}, \quad \ell = 1, \ldots, L, \tag{6}$$

which are defined in dependence of a common information criterion $IC$, such as $AIC$ or $BIC$ applied to each of the $L$ models, and prior model weights $p_\ell$. If each model is given the same prior model weight, the $p_\ell$ cancel out in Equation 6.

*Design considerations I: Power and sample size calculations*

An important step at the planning phase of any clinical trial is to properly design the study in order to achieve the study objectives. Because dose finding studies have two major goals, PoC testing and dose estimation, different criteria can be used to design a study. Dette *et al.* (2008) derived optimal designs, which minimize the (asymptotic) variance of the $MED$ estimate. Using their approach, asymptotic confidence intervals for the $MED$ can be calculated, conditional on a selected model. At the planning stage one would then specify the maximum width of the confidence interval and calculate the sample size necessary to ensure a certain precision of the $MED$ estimate.

An alternative approach is to focus on calculating the sample size necessary to achieve a pre-specified power to detect PoC (Pinheiro *et al.* 2006a). We thus start by introducing the power calculation under a given specific model $m$ from the candidate set $\mathcal{M}$, generalize it afterwards to multiple models and finally focus on sample size calculation.

The power of the MCP procedure is determined by the distribution of $T_{\max}$ under the alternative hypothesis that the $m$-th dose-response model is true. Under this assumption, the mean responses at the doses $d_1, \ldots, d_k$ are $\boldsymbol{\mu}_m = (f_m(d_1, \boldsymbol{\theta}_m), \ldots, f_m(d_k, \boldsymbol{\theta}_m))^\top$. The power to detect a dose-response signal (i.e., PoC) under model $m$ for sample sizes $\boldsymbol{n} = (n_1, \ldots, n_k)^\top$ is then

$$P(\max_l T_l \geq q_{1-\alpha} | \boldsymbol{\mu} = \boldsymbol{\mu}_m) = 1 - P(T_1 < q_{1-\alpha}, \ldots, T_M < q_{1-\alpha} | \boldsymbol{\mu} = \boldsymbol{\mu}_m). \tag{7}$$

It follows from the properties of the multivariate $t$ distribution and the assumptions in Equation 1, that, under the $m$-th model, the contrast test statistics $T_1, \ldots, T_M$ are jointly distributed as non-central multivariate $t$ with $N - k$ degrees of freedom and correlation matrix $\boldsymbol{R} = (\rho_{ij})$. The non-centrality parameter vector is $\boldsymbol{\delta}_m = (\delta_{m1}, \ldots, \delta_{mM})^\top$, where

$$\delta_{ml} = \frac{\sum_{i=1}^k c_{li} \mu_{mi}}{\sigma \sqrt{\sum_{i=1}^k c_{li}^2 / n_i}}, \; l = 1, \ldots, M.$$

Again, the **mvtnorm** package can be used to calculate the necessary probabilities.

So far we have only considered the power calculation under a single model $m$. In practice we would rather account for the inherent model uncertainty. To this end, we would calculate the power for each of the $M$ models from the candidate set $\mathcal{M}$ and aggregate the resulting values into a single *combined* measure of power, such as the (weighted) average, the minimum or a quantile. The sample size is then calculated as the smallest sample size ensuring a minimum *combined power* value, say $\pi^*$, to detect PoC under the assumed set of dose-response mean vectors. We restrict ourselves to the case that either the allocation weights $r_i \geq 0$, subject to $\sum_i r_i = 1$ or the allocation ratios $\rho_i$ relative to the dose group with the fewest patients, i.e., $\rho_i = r_i / \min(r_i)$ are prespecified. The group sample sizes $\boldsymbol{n} = (n_1, \ldots, n_k)$ can then be obtained from $n_i = N r_i$ for allocation weights or from $n_i = \rho_i n_{min}$, where $n_{min}$ is the smallest group sample size. Since the combined power is a monotone increasing function of

$N$ (or $n_{min}$, if allocation ratios are specified) a unique smallest integer giving a power larger than $\pi^*$ exists (see also Pinheiro *et al.* 2006a). The bisection search method can be used to obtain the sample size ensuring a pre-specified combined power $\pi^*$. In practice, rounding techniques need to be applied to obtain integer sample sizes.

### *Design considerations II: Sensitivity analysis*

In the derivations above we conditioned on the mean vectors $\boldsymbol{\mu}_m$ in Equation 7 and hence on the parameters $\boldsymbol{\theta}_m = (\theta_{m0}, \theta_{m1}, \boldsymbol{\theta}_m^*)^\top$. Since the sample size is calculated under this condition, it is critical that the model parameters are reliably determined. For the determination of location and scale parameters $\theta_{m0}$ and $\theta_{m1}$, prior knowledge about the expected placebo response $\delta_0$ and the maximum response $\delta_{\max}$ can be used at the design stage. It is typically straightforward to plug in these quantities into the model equations, assuming that $\boldsymbol{\theta}_m^*$ is known and then solving for $\theta_{m0}$ and $\theta_{m1}$, see Pinheiro *et al.* (2006a) for a more detailed description of this approach.

Based on prior knowledge about the shape of the model function, Pinheiro *et al.* (2006b) discussed strategies to obtain guesstimates for the standardized model parameters $\boldsymbol{\theta}_m^*$. The elicitation of prior information for $\boldsymbol{\theta}_m^*$ may impact both the design *and* the analysis of a dose finding study using the MCP-Mod methodology, as the guesstimates are used to obtain the optimal model contrasts at the MCP step, which in turn determine the effective power to detect PoC. Therefore, it is of importance to investigate the sensitivity of the procedure to misspecification of the parameters in the standardized models and, in particular, the impact it has on the effective power to detect PoC. Pinheiro *et al.* (2006a) considered different measures of loss in power associated with a misspecification of the standardized model parameters. One possibility, subsequently denoted as $LP_1$, is to calculate the difference between the nominal power (the power obtained, when the guesstimate is correct) and the actual power (the power obtained, when the used guesstimate does not coincide with the true parameter), i.e.,

$$LP_1 = \text{nominal power} - \text{actual power}. \tag{8}$$

Thus, $LP_1$ can be interpreted as the difference between the power that was intended for the study and the power one actually obtains. Alternatively, one could also calculate the difference between the power that could be achieved if the true parameter values were known at the design stage (potential power) and the actual power. This is denoted by $LP_2$ and hence

$$LP_2 = \text{potential power} - \text{actual power}.$$

Graphical methods can be used to display the loss in power for a range of true standardized model parameters. From our experience the loss in power associated with misspecification of the parameters in the standardized model function is often negligible for reasonable candidate sets, because dose-response models with parameter vectors $\boldsymbol{\theta}_m$ deviating from the guesstimate $\boldsymbol{\theta}_m^*$ are often detected from some other model in the candidate set. In cases where the loss in power is not acceptable, the inclusion of an additional model in the candidate set could be considered.

# 3. The MCPMod package

In this section we describe the R package **MCPMod** for implementing the MCP-Mod methodology. The package consists of two main parts (see also Figure 2). The first part contains several functions that are useful for planning a trial: calculation of the optimal contrasts and the critical value (`planMM`), the sample size (`sampSize`) or functions that support the selection of a 'good' candidate set and sensitivity analysis (`guesst`, `plotModels`, `powerMM`, `LP`). The second part consists of one main function named `MCPMod` that implements the full MCP-Mod approach for analysis of a given dose-response data set.

## 3.1. Preliminaries

Before illustrating the different functions in more detail we first describe how to specify the candidate set of models $\mathcal{M}$ for these functions. Table 1 gives an overview of the dose-response models that are implemented (note that user-defined non-linear models can also be specified, see the package documentation for details). The candidate set of models needs to be specified as a `list`, where the list elements should be named according to the underlying dose-response model function (see Table 1) and the individual list entries should correspond to the required guesstimates or `NULL` if no guesstimates are needed. Suppose, for example, we want to include in our candidate set a linear model, an $E_{max}$ model and a logistic model. From the standardized model functions in Table 1 we see that we need to specify one guesstimate for the $E_{max}$ model ($ED_{50}$ parameter), two guesstimates for the logistic model ($ED_{50}$ and $\delta$) and none for the linear model (since its standardized model function does not contain any unknown parameters). Suppose our guesstimate for the $ED_{50}$ parameter of the $E_{max}$ model is 0.2, while the guesstimate for $(ED_{50}, \delta)^{\top}$ for the logistic model is $(0.25, 0.09)^{\top}$. We then specify the list

```
R> mods1 <- list(linear = NULL, emax = 0.2, logistic = c(0.25, 0.09))
```

In some cases one might want to include several model shapes per model class. For example, if the candidate model set includes two $E_{max}$ model shapes, two logistic model shapes, a beta
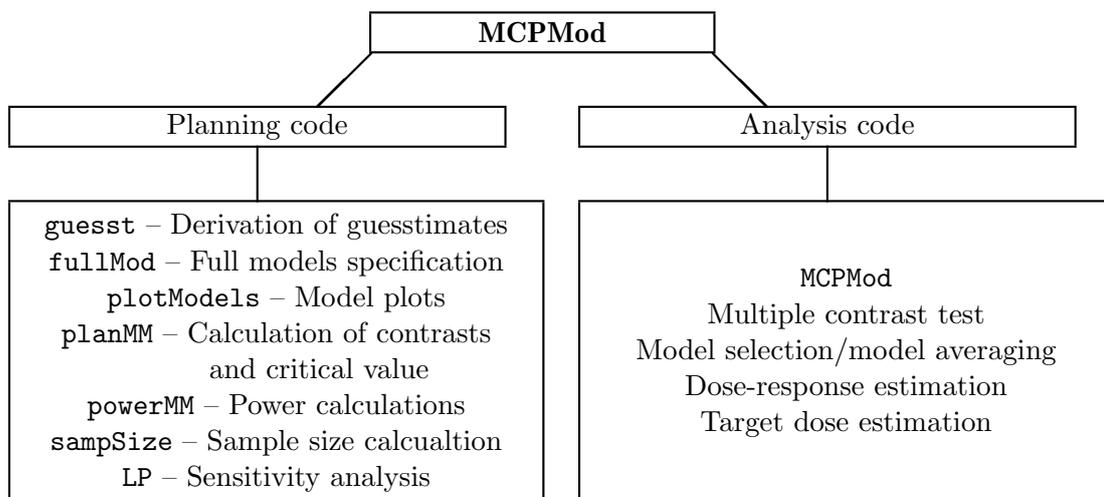
Figure 2: Overview of main functions in the **MCPMod** package.

| Name | $f(d, \boldsymbol{\theta})$ | $f^0(d, \boldsymbol{\theta}^*)$ | (*) | (♯) |
|---|---|---|---|---|
| `linear` | $E_0 + \delta d$ | $d$ | | |
| `linlog` | $E_0 + \delta \log(d + c)$ | $\log(d + c)$ | | $c$ |
| `quadratic` | $E_0 + \beta_1 d + \beta_2 d^2$ | $d + \delta d^2$ if $\beta_2 < 0$ | $\delta$ | |
| `emax` | $E_0 + E_{\max} d / (ED_{50} + d)$ | $d / (ED_{50} + d)$ | $ED_{50}$ | |
| `logistic` | $E_0 + E_{\max} / \left\{ 1 + \exp\left[ (ED_{50} - d) / \delta \right] \right\}$ | $1 / \left\{ 1 + \exp\left[ (ED_{50} - d) / \delta \right] \right\}$ | $(ED_{50}, \delta)^\top$ | |
| `exponential` | $E_0 + E_1(\exp(d/\delta) - 1)$ | $\exp(d/\delta) - 1$ | $\delta$ | |
| `sigEmax` | $E_0 + E_{\max} d^h / (ED_{50}^h + d^h)$ | $d^h / (ED_{50}^h + d^h)$ | $(ED_{50}, h)^\top$ | |
| `betaMod` | $E_0 + E_{\max} B(\delta_1, \delta_2)(d/D)^{\delta_1}(1 - d/D)^{\delta_2}$ | $B(\delta_1, \delta_2)(d/D)^{\delta_1}(1 - d/D)^{\delta_2}$ | $(\delta_1, \delta_2)^\top$ | $D$ |

Table 1: Dose-response models implemented in the **MCPMod** package. Column (*) lists for each model the parameters for which guesstimates are required and the order in which they need to be specified in the `models` list, while column (♯) lists the parameters, which fixed and not estimated. For the beta model $B(\delta_1, \delta_2) = (\delta_1 + \delta_2)^{\delta_1 + \delta_2} / (\delta_1^{\delta_1} \delta_2^{\delta_2})$ and for the quadratic model $\delta = \frac{\beta_2}{|\beta_1|}$. For the quadratic model the standardized model function is given for the concave-shaped form.

model shape and a linear model shape the model list would look like

```
R> mods2 <- list(linear = NULL, emax = c(0.05, 0.2), betaMod = c(0.5, 1),
+    logistic = matrix(c(0.25, 0.7, 0.09, 0.06), byrow = FALSE, nrow = 2))
```

Thus, if multiple model shapes from the same model class are to be used, the parameters are handed over as a matrix, for models having two parameters in the standardized model function, and as a vector for one-parameter standardized models. This general structure applies to all built-in models. Note that the linear-in-log and the beta models also contain a parameter ($c$ and $D$, respectively, see Table 1) that is not estimated from the data but needs to be pre-specified. These parameters are not handed over via the candidate model list but via seperate arguments `scal` (corresponding to $D$) and `off` (corresponding to $c$) respectively to the top-level functions.

## 3.2. Planning code

In this section we provide a brief overview of the functions `guesst`, `plotModels`, `fullMod`, `planMM`, `powerMM`, `sampSize` and `LP`. These functions are useful for designing a trial using MCP-Mod. For a detailed description of the arguments to the functions we refer to the documentation of the package.

### *Function* `guesst`

The selection of suitable guesstimates and model shapes is a major aspect of the MCP-Mod methodology. Incorporating contrasts/models that are likely to be true (and excluding those that are very unlikely) can greatly improve the power of the methodology. The `guesst` function supports the translation of clinical knowledge available prior to the start of a study into the required guesstimates. The function calculates the guesstimates according to the percentage $p^*$ of the maximum effect that is achieved at a certain dose $d^*$. Suppose, for example, we want to calculate a guesstimate for the $ED_{50}$ parameter from the $E_{\max}$ model. If we expect a response of 90% at dose 0.2, the $ED_{50}$ guesstimate can be calculated by calling

```
R> guesst(d = 0.2, p = 0.9, model = "emax")

      ed50
0.02222222
```

With this guesstimate, the standardized model for the $E_{\max}$ model is given by $f^0(d, ED_{50}) = d/(0.022222+d)$, and the optimal contrast can be calculated from Equation 3. For models with two standardized model parameter one $(d^*, p^*)$ pair is not sufficient to obtain guesstimates for the standardized model parameters. For example, for the logistic model we need to specify two pairs to obtain a guesstimate

```
R> guesst(d = c(0.05, 0.2), p = c(0.2, 0.9), model = "logistic")

      ed50      delta
0.1080279 0.0418583
```

In this example the standardized model function for the logistic model is given by $f^0(d, ED_{50}, \delta) = 1/\{1 + \exp[(0.1080279 - d)/0.0418583]\}$, from which the corresponding optimal contrast can be obtained. In a similar way one can obtain guesstimates with the `guesst` function for all built-in models.

### *Function* `plotModels`

Before deciding for any particular candidate set of model shapes it is useful to display them graphically. This can be done with the `plotModels` function. Since the model shapes, specified in the models list, do not depend on the location (defined through the baseline effect) and scale (defined through the maximum effect) of the model, one additionally needs to specify those via the `base` and `maxEff` arguments. Using the candidate set `mods2` defined above (and setting the `scal` parameter of the beta model equal to 1.2), a graphical representation can be obtained as follows (see Figure 3 for the output)

```
R> doses <- c(0, 0.05, 0.2, 0.6, 1)
R> plotModels(mods2, doses, base = 0, maxEff = 0.4, scal = 1.2)
```

### *Function* `fullMod`

Similar to the `plotModels` function above, also other functions (`powerMM`, `sampSize`, `LP`) require information about the doses, the full model functions, i.e., the candidate model shapes, the baseline effect, the maximum effect and possible other additional parameters like `off` or `scal`. The `fullMod` function derives the full model functions (i.e., the location and scale parameters) for each model from the stated information (see Section 2.2) and packages this with the used dose levels into a `fullMod` object, which can then be used as an input parameter for the four above mentioned functions. When assuming the baseline effect 0 and the maximum effect 0.4 and using the candidate set `mods2` (and setting the `scal` parameter of the beta model equal to 1.2) one can package this information via

```
R> doses <- c(0, 0.05, 0.2, 0.6, 1)
R> fmods2 <- fullMod(mods2, doses, base = 0, maxEff = 0.4, scal = 1.2)
```
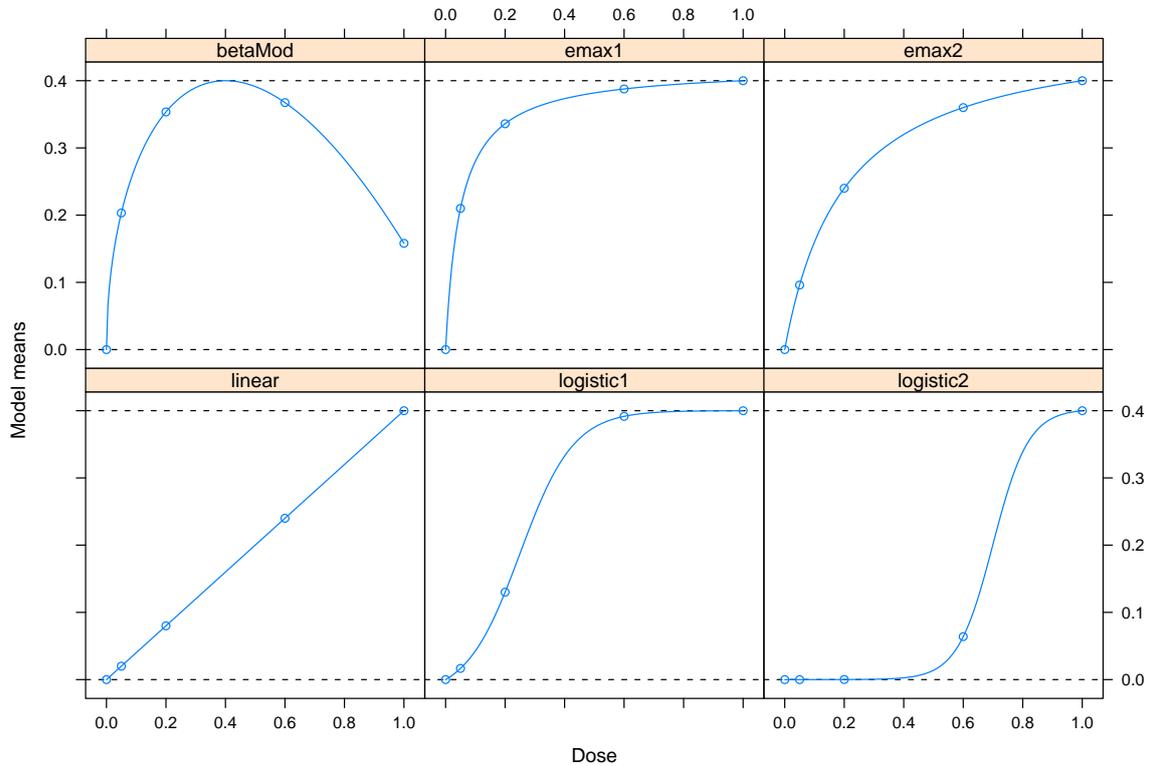
Figure 3: Model shapes for the selected candidate model set, produced with `plotModels` function.

### *Function* `planMM`

The `planMM` function calculates the quantities necessary to conduct the multiple contrast test: The optimal model contrasts and their correlations (see Equations 3 and 4) and the critical value using the **mvtnorm** package. This information is returned in a `planMM` object. The arguments `alpha` and `twoSide` determine the significance level and sidedness of the test. By default one-sided testing at level $\alpha = 0.025$ is performed. The sample size allocations are handed over as a vector via the `n` argument (for balanced allocations a single number is sufficient). Assuming a balanced allocation of 20 patients per dose group, the candidate set `mods2` and the doses from above, the `planMM` function can be called as follows

```
R> pM <- planMM(mods2, doses, n = 20, alpha = 0.05, twoSide = FALSE,
+     scal = 1.2)
R> pM

MCPMod planMM

Optimal Contrasts:
     linear  emax1   emax2 betaMod logistic1 logistic2
0    -0.437 -0.799  -0.643  -0.714    -0.478    -0.267
```

```
0.05 -0.378 -0.170 -0.361  -0.043    -0.435    -0.267
0.2  -0.201  0.207  0.061   0.452    -0.147    -0.267
0.6   0.271  0.362  0.413   0.498     0.519    -0.083
1     0.743  0.399  0.530  -0.192     0.540     0.883

Critical Value (alpha = 0.05, one-sided): 2.139

Contrast Correlation Matrix:
          linear emax1 emax2 betaMod logistic1 logistic2
linear     1.000 0.766 0.912   0.229     0.945     0.905
emax1      0.766 1.000 0.949   0.774     0.828     0.525
emax2      0.912 0.949 1.000   0.606     0.956     0.686
betaMod    0.229 0.774 0.606   1.000     0.448    -0.130
logistic1  0.945 0.828 0.956   0.448     1.000     0.717
logistic2  0.905 0.525 0.686  -0.130     0.717     1.000
```

The first part of the output shows the optimal contrast coefficients for the different models. The representation of the optimal contrast is unique as we imposed the condition of unit Euclidean length. In the output we then obtain the multiplicity adjusted critical value for the maximum contrast and finally the correlations of the contrasts. In this example some contrasts are quite highly correlated. For example, the correlation between emax2 and `logistic1` is `0.956`, indicating that both describe similar dose-response shapes, as can also be seen in
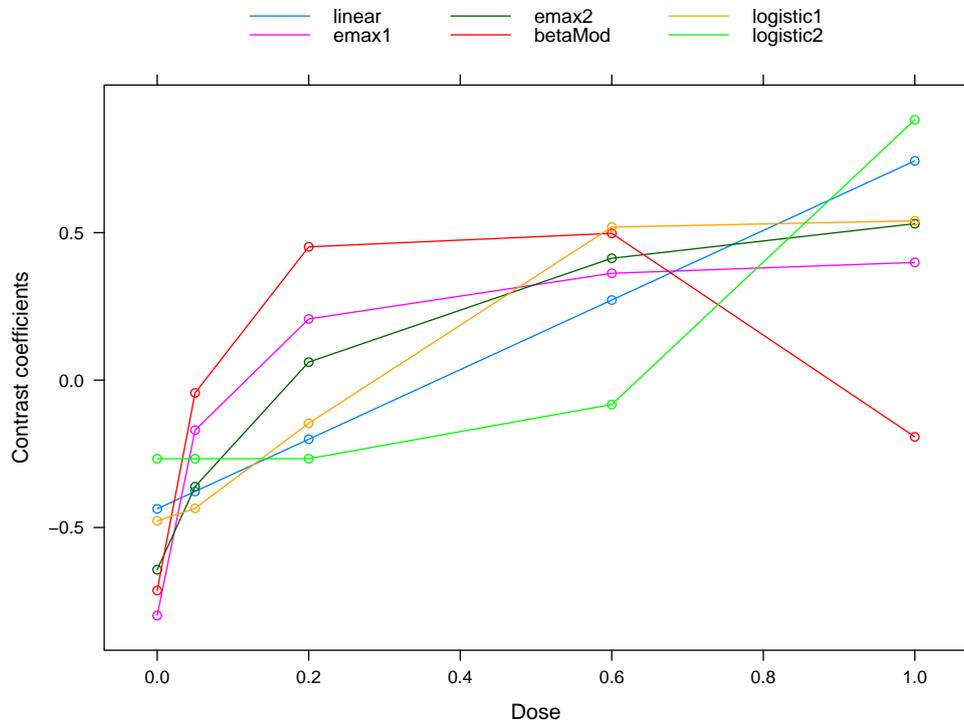


Figure 4: Graphical display of optimal contrasts.

Figure 3. The beta model contrast however, seems to be relatively different from the others. This is due to the fact that the beta model shape is, contrary to the other model shapes, not monotone. The contrasts can also be graphically displayed using the `plot` method for `planMM` objects, e.g., `plot(pM)` (see Figure 4).

*Function* `powerMM`

The `powerMM` function is designed to calculate the power to detect the model shapes in the candidate set for different sample sizes. We need to hand over either an object of class `fullMod` or the doses, the baseline and the maximum effect via `doses`, `base` and `maxEff` and the standard deviation of the response via `sigma`. One can calculate the power for sample sizes ranging from `lower` to `upper` in stepsizes `step`. Summary functions can be used to combine the different power values for the different model shapes into one value, as described in Section 2.2. By default the minimum, the mean and the maximum power are calculated. The resulting power values are returned as an object of class `powerMM` in a matrix. There exists also a `plot` method to display the results graphically. Using the information packaged in the `fmods2` object from above one obtains the following result

```
R> pM <- powerMM(fmods2, sigma = 1, alpha = 0.05, lower = 10,
+    upper = 110, step = 10)
R> plot(pM, line.at = 0.9, models = "none")
```
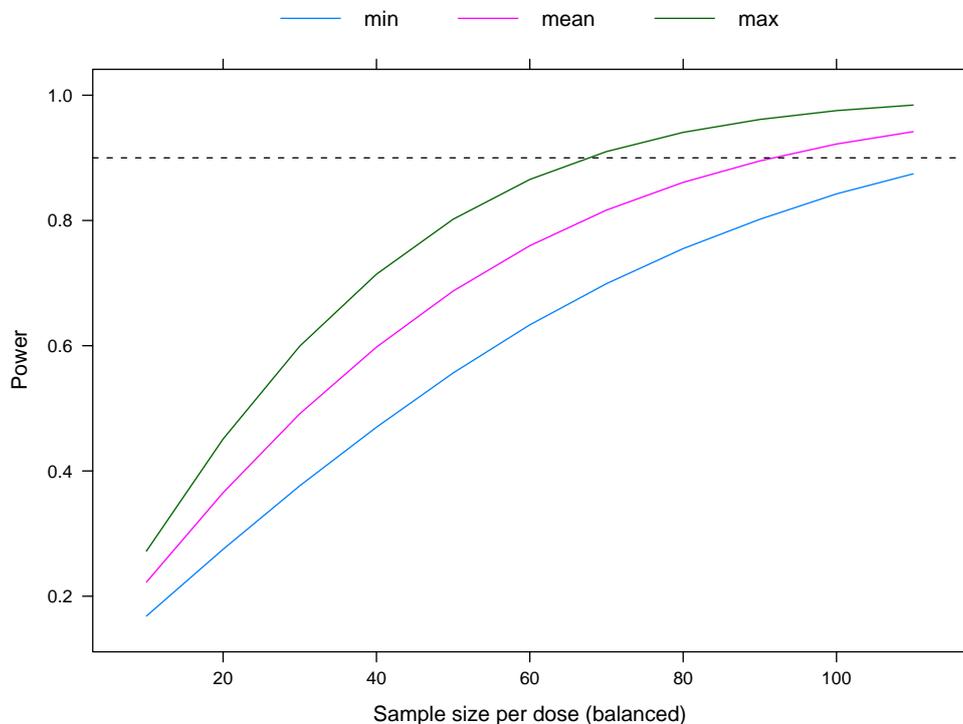


Figure 5: Power to detect PoC under the assumed candidate set for different summary functions.

In Figure 5 it can be seen that a mean power of 90 % is achieved with approximately 90 patients per dose. Note that the power can also be calculated for unbalanced but fixed allocations. The allocation ratios (or allocation weights, depending on the value of the `typeN` argument) then need to be supplied via the `alRatio` argument. In the plot above only the summary power values are displayed, although the `plot` method for `powerMM` allows the display of the power values for the individual candidate models as well.

### Function `sampSize`

The `sampSize` function calculates the necessary sample size to achieve a pre-specified *combined* power value. As input parameters we need a `fullMod` object (or manually `doses`, `base`, `maxEff`) and `sigma`. Together with the candidate set, these parameters form the 'alternatives' for which the power is calculated. A summary function (via `sumFct`) to combine the individual power values into one value and the power level we want to achieve (via `power`) need to be provided as well. For the bisection search algorithm an upper bound for the target sample size (via `upperN`) needs to be provided as a starting value. The starting value for the lower bound needed for the bisection is derived internally as `upperN/2`, but can also be handed over manually via `lowerN`. When the starting values for the upper and lower bound do not bracket a solution the bounds are extended automatically. For the information packaged in the `fmods2` object the result is as follows

```
R> sampSize(fmods2, sigma = 1, sumFct = mean, power = 0.9, alpha = 0.05,
+    twoSide = FALSE, upperN = 100)


MCPMod sampSize

Input parameters:
 Summary Function: mean
 Desired combined power value: 0.9
 Level of significance: 0.05 (one-sided)
 Allocations: balanced

Sample size per group: 92

Associated mean power: 0.9013
Power under models:
   linear     emax1     emax2   betaMod logistic1 logistic2
   0.9106    0.8997    0.9163    0.8110    0.9647    0.9058
```

As seen from the output, the `sampSize` function returns the desired group sample size and the associated combined power. In our example we thus need 92 patients per group to guarantee a mean power of 90%. The `sampSize` function also returns the individual power values under the different models in the candidate set. Note that in the example above we assumed a balanced sample size allocation. Fixed allocation proportions can be specified via the `alRatio` argument. If `typeN = "arm"`, the code assumes that allocation *ratios* are passed to `alRatio`, which means that the bisection search algorithm varies the sample size $n_{min}$ in the dose group with the fewest number of patients, and returns the smallest $n_{min}$ such that

the combined power is larger than `power`. If `typeN = "total"` allocation *weights* are assumed and the overall sample size $N$ is iterated.

*Function* `LP`

The `LP` function is designed to calculate the loss in power associated with misspecification of the guesstimates for one model in the candidate set. To illustrate the function we use a very simple candidate set consisting of only a linear and an $E_{max}$ shape and illustrate the calculation of $LP_1$ (see Equation 8). We select 0.15 as the guesstimate for the $ED_{50}$ parameter, and want to investigate the loss in power in the interval $[0.03, 0.8]$ (specified via `paramRange`). Hence we calculate how much power we loose, if an alternative $ED_{50}$ value is true, but we selected 0.15 as our guesstimate. As before `doses`, `base`, `maxEff` (or an object of class `fullMod`) and `sigma` need to be specified together with the sample size. After calling the `LP` function we display the results using the associated `plot` method. The optional `spldf` argument determines the degrees of freedom for the spline that is used to smooth the power values in the plot.

```
R> mods3 <- list(linear = NULL, emax = 0.15)
R> Lfit <- LP(mods3, model = "emax", type = "LP1",
+    paramRange = c(0.03, 0.8), len = 30, doses = doses, n = 92,
+    base = 0, maxEff = 0.4, sigma = 1, alpha = 0.05, twoSide = FALSE)
R> plot(Lfit, spldf = 25)
```
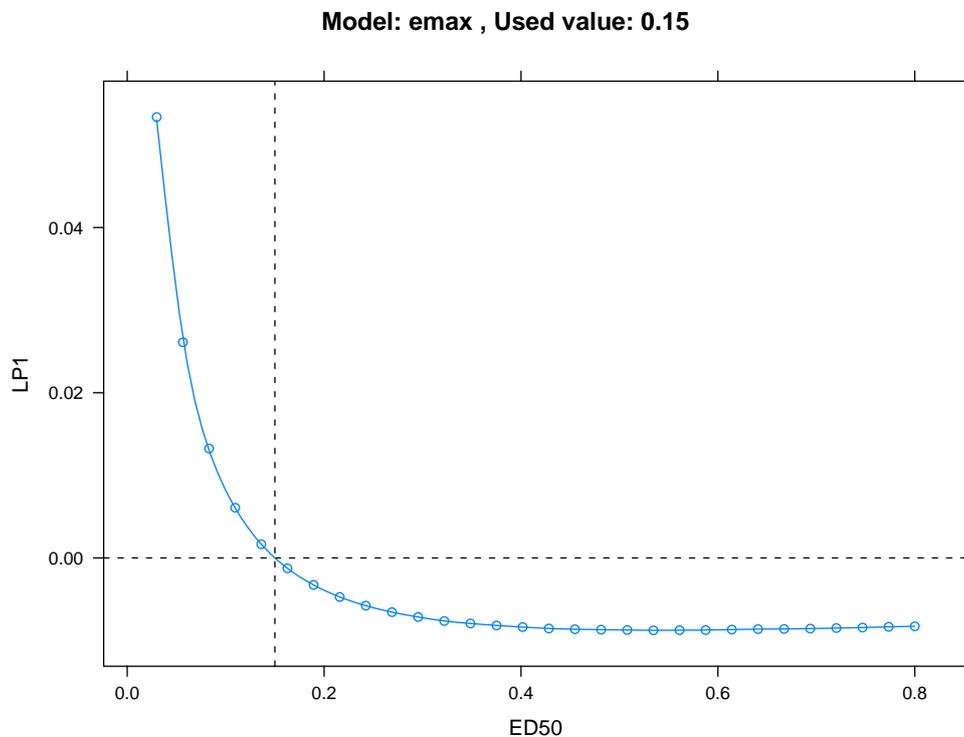


Figure 6: Difference of actual and nominal power for Emax model.

As seen from Figure 6 the loss in power is relatively large if a small $ED_{50}$ value is true. If the true $ED_{50}$ is equal to the guesstimate then the actual power and nominal power coincide. For $ED_{50}$ larger than the specified guesstimate we actually gain power, because the $E_{\max}$ model becomes almost linear and is captured by the linear model included in the candidate set.

## 3.3. Analysis code

The analysis functionalities are incorporated in one main function `MCPMod`, which implements the full MCP-Mod approach. According to the methodology described in Section 2, it consists of two main steps: (1) MCP-step (calculation of optimal contrasts, critical value, contrast test statistics and possibly $p-$values and selection of the set of significant models) and (2) modeling step (model fitting, model selection/model averaging and dose estimation).

We now describe some of the more important arguments for the `MCPMod` function. For a complete description of the `MCPMod` function we refer to the online documentation. The dose-response data set is handed over to the `MCPMod` function via the `data` argument. It should be handed over as a `data frame` containing two columns corresponding to the dose levels and the response values. The `selModel` argument determines how to select a dose estimation model out of the set of significant models (if there are any significant models). One can choose between the maximum contrast test statistic (the default option), the AIC, the BIC or model averaging based on either the AIC or the BIC (see Section 2.2). Another important argument is `doseEst`, which determines the dose estimator to be used. Three slightly different estimators for the $MED$ are currently implemented (see Bretz *et al.* (2005) for a detailed description of those three estimators, option `"MED2"` is the default value, corresponding to the estimator described in Section 2.2) as well as an estimator of the $EDp$. Additional parameters for the dose estimators (such as $\gamma$ for $MED$ estimators (default: $\gamma = 0.1$) and $p$ for the $ED$ estimator (default: $p = 0.5$)) are handed over via the `dePar` argument. The clinical relevance threshold $\Delta$ is handed over via the `clinRel` argument. The `pVal` argument determines, whether multiplicity adjusted $p-$values for the multiple contrast test should be calculated or not (per default p-values are not calculated).

To illustrate the `MCPMod` function we use the dose-response data set `biom` used by Bretz *et al.* (2005) to illustrate the MCP-Mod methodology. The data result from a randomized double-blind parallel group trial with a total of 100 patients being allocated to either placebo or one of four active doses coded as 0.05, 0.20, 0.60, and 1, with 20 patients per group. Here, we use the `MED2` estimator with $\gamma = 0.05$ to estimate the $MED$, the clinical threshold $\Delta$ is set to 0.4 and the dose estimation model is selected according to the maximum contrast test statistic. Employing the candidate model set `mods2` the results can be obtained by calling

```
R> data("biom")
R> dfe <- MCPMod(biom, mods2, alpha = 0.05, dePar = 0.05, pVal = TRUE,
+    selModel = "maxT", doseEst = "MED2", clinRel = 0.4, scal = 1.2)
```

A brief summary of the results is available via the `print` method for `MCPMod` objects

```
R> dfe

MCPMod
```

```
PoC (alpha = 0.05, one-sided): yes
Model with highest t-statistic: emax2
Model used for dose estimation: emax
Dose estimate:
MED2,90%
     0.17
```

From the output we conclude that the maximum contrast is significant at one-sided level 0.05. Thus a significant dose-response relationship can be established, i.e., positive PoC. Furthermore we conclude that `emax2` has the largest test statistic among all contrasts and consequently the $E_{max}$ model was used for the dose-estimation step. The $MED$ estimate is 0.17. The 90% in the $MED$ estimate refers to the confidence level of $L(d)$ used in the dose estimator (see Section 2.2). A more detailed summary of the results is available via the `summary` method.

```
R> summary(dfe)


MCPMod


Input parameters:
 alpha = 0.05 (one-sided)
 model selection: maxT
 clinical relevance = 0.4
 dose estimator: MED2 (gamma = 0.05)


Optimal Contrasts:
      linear   emax1   emax2 betaMod logistic1 logistic2
0     -0.437 -0.799 -0.643  -0.714    -0.478    -0.267
0.05  -0.378 -0.170 -0.361  -0.043    -0.435    -0.267
0.2   -0.201  0.207  0.061   0.452    -0.147    -0.267
0.6    0.271  0.362  0.413   0.498     0.519    -0.083
1      0.743  0.399  0.530  -0.192     0.540     0.883


Contrast Correlation:
          linear emax1 emax2 betaMod logistic1 logistic2
linear     1.000 0.766 0.912   0.229     0.945     0.905
emax1      0.766 1.000 0.949   0.774     0.828     0.525
emax2      0.912 0.949 1.000   0.606     0.956     0.686
betaMod    0.229 0.774 0.606   1.000     0.448    -0.130
logistic1  0.945 0.828 0.956   0.448     1.000     0.717
logistic2  0.905 0.525 0.686  -0.130     0.717     1.000


Multiple Contrast Test:
          Tvalue pValue
emax2      3.464  0.001
emax1      3.339  0.002
logistic1  3.235  0.002
```

```
linear     2.972  0.007
betaMod    2.402  0.028
logistic2  2.074  0.057


Critical value: 2.138



Selected for dose estimation:
  emax

Parameter estimates:
emax model:
   e0   eMax  ed50
0.322 0.746 0.142

Dose estimate
MED2,90%
    0.17
```

The summary output includes some information about important input parameters that were used when calling `MCPMod`. Then the output includes also the optimal contrasts and the contrast correlations together with the contrast test statistics, the multiplicity adjusted $p$−values and the critical value. Finally, information about the fitted dose-response model, its parameter estimates and the target dose estimate are displayed.

A graphical display of the dose-response model used for dose estimation can be obtained via the `plot` method for `MCPMod` objects (see Figure 7). When `complData = TRUE`, the full dose-response data set is plotted instead of only the group means. The `clinRel` option determines whether the clinical relevance threshold should be displayed.

```
R> plot(dfe, complData = TRUE, clinRel = TRUE)
```

To illustrate the different options available for the `MCPMod` function we will now re-analyze the `biom` data set with different input parameters. Specifically, we will now apply model averaging techniques. The target dose is hence estimated as the weighted average of the dose estimates under the different significant models. The weights are determined via the AIC criterion (see Equation 6) with uniform prior weights (which is the default). The target dose we are now interested in is the $ED95$, which is the dose that achieves 95 % percent of the maximum effect.

```
R> dfe2 <- MCPMod(biom, mods2, alpha = 0.05, dePar = 0.95,
+    selModel = "aveAIC", doseEst = "ED", scal = 1.2)
R> dfe2

MCPMod

PoC (alpha = 0.05, one-sided): yes
Model with highest t-statistic: emax2
```
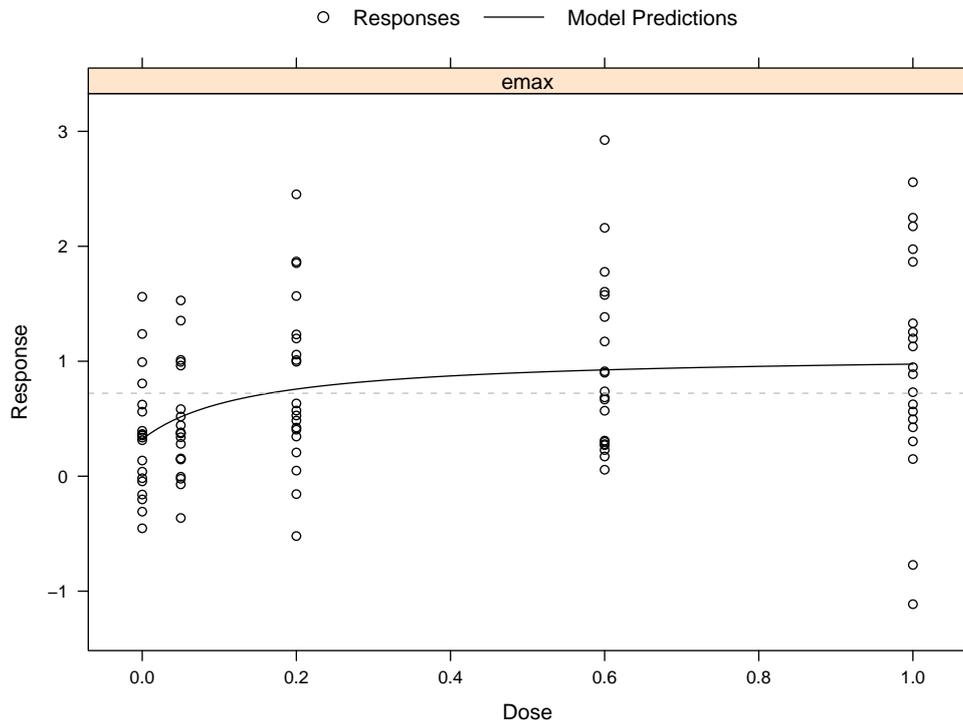
Figure 7: Fitted model with data set.

```
Models used for dose estimation: emax logistic linear betaMod
Dose estimate:
ED95%
0.669
```

The output of the `print` method now contains the four models selected for dose-response estimation as well as the model averaged $ED95$ estimate. We edited the output of the summary method here as there is some overlap with the previous call to the `summary` function.

```
R> summary(dfe2)

MCPMod

Input parameters:
 alpha = 0.05 (one-sided)
 model selection: aveAIC
 prior model weights:
     emax logistic   linear  betaMod
    0.25     0.25     0.25     0.25
 dose estimator: ED (p = 0.95)
.
.
```

```
.
AIC criterion:
    emax logistic    linear  betaMod
  219.14    220.83    220.50    221.32

Selected for dose estimation:
  emax logistic linear betaMod


Model weights:
    emax logistic    linear  betaMod
   0.440     0.189     0.223     0.148

Parameter estimates:
emax model:
   e0   eMax   ed50
0.322 0.746 0.142
logistic model:
   e0   eMax   ed50 delta
0.169 0.773 0.087 0.071
linear model:
(Intercept)         dose
      0.492        0.559
betaMod model:
    e0    eMax delta1 delta2
 0.329   0.669   0.573   0.321

Dose estimate
Estimates for models
      emax logistic linear betaMod
ED95% 0.71      0.32   0.95    0.57
Model averaged dose estimate
ED95%
0.669
```

In addition to the results already described in the `summary(dfe)` call, the output now also contains information about the AIC of the different models and the model weights. All model fits are given and the $ED95$ estimate obtained for all models, as well as the model weighted average of the dose estimates.

A graphical display of the fitted model functions can be obtained via the `plot` method, here we just plot the model means but also include the estimated ED95 in the plot (see Figure 8).

```
R> plot(dfe2, doseEst = TRUE)
```
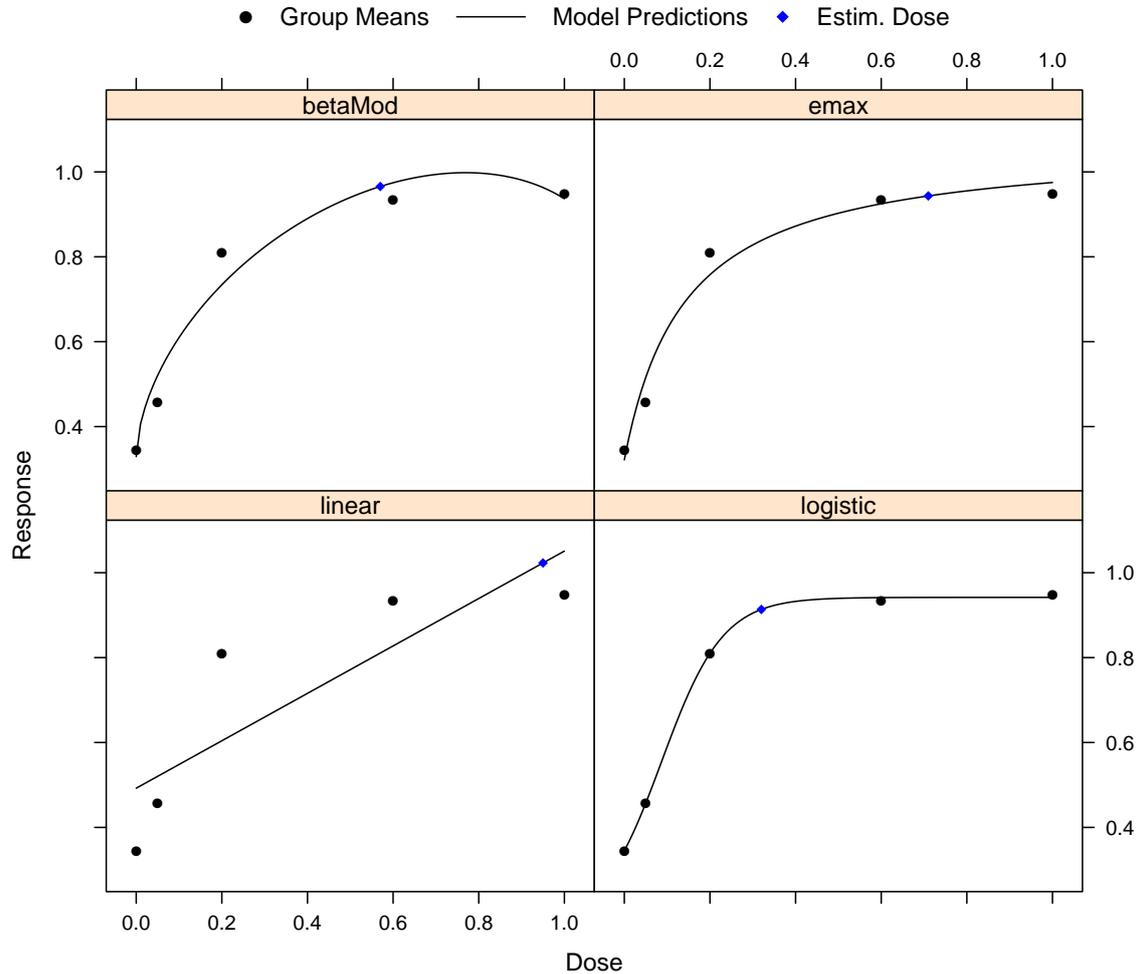
Figure 8: Fitted models with data set.

# 4. Summary and outlook

In this paper we have reviewed the MCP-Mod methodology including its most recent developments and introduced the **MCPMod** package. The paper is based on version 1.0-1 of the package, but the package will stay under development. Future versions will, among other features, include bootstrap methods for calculating confidence intervals on target dose estimates and the fitted model function, inclusions of covariates as well as a version of the Golub-Pereyra algorithm, which allows for box constraints. Updated versions of this document, reflecting potential changes in the package can be found as a vignette enclosed in the **MCPMod** package.

# Acknowledgments

# References

Bornkamp B (2006). *Comparison of Model-Based and Model-Free Approaches for the Analysis of Dose-Response Studies.* Master's thesis, Fachbereich Statistik, Universität Dortmund. URL http://www.statistik.tu-dortmund.de/~bornkamp/diplom.pdf.

Bretz F, Pinheiro JC, Branson M (2005). "Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies." *Biometrics*, **61**, 738–748.

Buckland ST, Burnham KP, Augustin NH (1997). "Model Selection an Integral Part of Inference." *Biometrics*, **53**, 603–618.

Casella G, Berger RL (1990). *Statistical Inference.* Duxbury Press, Belmont, Calif.

Dette H, Bretz F, Pepelyshev A, Pinheiro JC (2008). "Optimal Designs for Dose Finding Studies." *Journal of the American Statisical Association*, **103**, 1225–1237.

Genz A, Bretz F (2002). "Methods for the Computation of Multivariate *t*-Probabilities." *Journal of Computational and Graphical Statistics*, **11**, 950–971.

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2009). ***mvtnorm****: Multivariate Normal and t Distributions.* R package version 0.9-4, URL http://CRAN.R-project.org/package=mvtnorm.

Golub G, Pereyra V (2003). "Separable Nonlinear Seast Squares: The Variable Projection Method and Its Applications." *Inverse Problems*, **19**, R1–R26.

Pinheiro JC, Bornkamp B, Bretz F (2006a). "Design and Analysis of Dose Finding Studies Combining Multiple Comparisons and Modeling Procedures." *Journal of Biopharmaceutical Statistics*, **16**, 639–656.

Pinheiro JC, Bretz F, Branson M (2006b). "Analysis of Dose-Response Studies – Modeling Approaches." In N Ting (ed.), "Dose Finding in Drug Development," pp. 146–171. Springer-Verlag, New York.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ruberg SJ (1995). "Dose Response Studies. I. Some Design Considerations." *Journal of Biopharmaceutical Statistics*, **5**(1), 1–14.

Tukey JW, Ciminera JL, Heyse JF (1985). "Testing the Statistical Certainty of a Response to Increasing Doses of a Drug." *Biometrics*, **41**, 295–301.

**Affiliation:**

Björn Bornkamp
Fakultät Statistik
Technische Universität Dortmund
DE-44221 Dortmund, Germany
E-mail: bornkamp@statistik.tu-dortmund.de

José Pinheiro
Biostatistics Department
Novartis Pharmaceuticals Corp.
East Hanover, New Jersey 07936-1080, United States of America
E-mail: jose.pinheiro@novartis.com

Frank Bretz
Biostatistics Department
Novartis Pharma AG
CH-4002 Basel, Switzerland
E-mail: frank.bretz@novartis.com