NOVARTIS

Development IIS

Efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty

Request for CHMP Qualification Opinion

Author(s): Björn Bornkamp, Frank Bretz, José Pinheiro

Document type: Request for Qualification Opinion

Document status: Final

Release date: 22 April 2013

Table of contents

List of tables

List of figures

1 Executive summary

Selection of a dose for confirmatory Phase III trials and potential market authorization is among the most difficult decisions in the whole development process and poor dose selection for confirmatory trials has persisted as a problem in the last decade. One of the main reasons for this is an inappropriate knowledge of the dose response relationship (efficacy and safety). In this document the MCP-Mod approach for dose response testing and estimation is presented that is intended to enable more informative Phase II study designs to provide a more solid basis for all subsequent dose selection strategies and decisions (Pinheiro et al., 2010).

Historically, dose finding studies were designed and analyzed based on multiple pairwise comparisons of the active doses against placebo. Such traditional methods are mostly performed within the analysis-of-variance hypothesis testing paradigm and known to be inefficient in the sense that they fail to fully incorporate the available information across the doses (Bornkamp et al., 2007). Modeling approaches assume a functional relationship between response and dose, taken as a quantitative factor, according to a certain parametric model, such as the Emax model. The fitted model is then used to describe the dose response relationship and estimate target doses of interest. Such an approach will often lead to more precise estimates of the dose response relationship, due to the interpolation of information across doses. In addition, it also provides flexibility in investigating the effect of doses not used in the actual study. However, the validity of its conclusions will highly depend on an appropriate choice of the dose response model, which in practice is often unknown. The MCP-Mod approach considered in this dossier is an approach for dealing with this model-uncertainty by combining principles of multiple comparisons with modeling techniques to overcome some of the shortcomings of applying either approach alone (Bretz et al., 2005).

The objective of the current submission is to seek qualification of the MCP-Mod approach, as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. The MCP-Mod approach is efficient in the sense that it uses the available data better than traditional pairwise comparisons. The MCP-Mod approach impacts both the design and the analysis of dose finding studies; see Figure 3-1 for details. At the trial design stage, a suitable set of candidate models is identified in repeated clinical team discussions, which also impacts decisions on the number doses, required sample sizes, patient allocations, etc. At the trial analysis stage, dose response is tested using suitable trend tests deduced from the set of candidate models. Once a dose response signal is established, the best model out of the pre-specified candidate model set is then used for dose response and target dose estimation.

A thorough review of the statistical and medical literature is presented in the current submission in support of the advantages of the methodology. In addition, we provide the results of extensive simulations studies to support the intended claim. In the past years, MCP-Mod has also been presented to and discussed with several major regulatory agencies, including EMA (Modeling & Simulation ITN Briefing Meeting on “Dose Finding Under Model Uncertainty” on July 15, 2009 in London; EMEA/457334/2009), FDA (1-day course to the CDER/OTS Office of Biostatistics on “Dose Finding Studies: Methods and Implementation” on August 12, 2008) and PMDA (half-day seminar on “Improving Dose Finding Studies: Principles, Methods, and Regulatory Perspectives” on December 11, 2009).

2 Statement of the need for and impact of the proposed novel methodology in clinical drug development

2.1 Statement of the need

A well-known problem of failed Phase III programs is often poor dose selection resulting from inappropriate knowledge of dose response relationship (efficacy and safety) at the end of the learning phase of drug development, i.e., Phase II. Selection of a dose (or doses) to carry into confirmatory Phase III trials is among the most difficult decisions in drug development. Although the exact numbers are unknown, it is believed that the high attrition rate plaguing the pharmaceutical industry in Phase III studies are due, at least in part, to inadequate dose selection for confirming safety and efficacy in the intended patient population – doses that are too low to achieve adequate benefit, as well as doses that are too high and lead to dose-related safety events. There is also evidence that, even after registration, dose adjustments in the label continue to be required with some frequency (Cross et al. 2002; Heerdink et al. 2002). In a broader context than just dose finding, Bayer Healthcare recently reviewed 67 of their in-house projects at replicating the findings in published research and reported that less than 1/4 were viewed as having been essentially replicated and over 2/3 had major inconsistencies leading to project termination (Prinz et al., 2011).

In recognition of this problem, the Pharmaceutical Innovation Steering Committee (PISC) of the Pharmaceutical Research and Manufacturers of America (PhRMA) formed in the spring of 2005 several working groups to look into different drivers of the decreasing success rates observed in drug development programs across the pharmaceutical industry, identified in a previous survey conducted by a consulting group. Among those was the “Adaptive Dose-Ranging Studies” (ADRS) working group. The objectives of this group were to develop new and evaluate novel adaptive and non-adaptive dose-ranging methods (one of them being MCP-Mod), and to provide methodological recommendations for industry and regulatory agencies on their use in clinical drug development. Details can be found in the white papers from the ADRS working group (Bornkamp et al. 2007; Pinheiro et al. 2010).

The basic difficulty in getting the right dose is the trade-off between wanted and un-wanted effects. A prerequisite for informed decision making and dose selection at end of Phase II is hence a solid characterization of the dose-response relationships. In the past, dose finding studies were often designed using a small number of doses and a narrow dose-range, often focused on the upper end of the dose response relationship. Only in recent years there was a noticeable shift towards investigating the full dose response relationship and estimating the so-called minimum effective dose (MED). The MED denotes the smallest dose achieving a pre-specified clinical treatment effect. Knowing the MED is important, because it defines a lower bound for therapeutically useful doses and most drugs have increasing potential for safety problems as dose is increased.

Figure 2-1 displays a non-exhaustive set of dose response profiles that are often seen in clinical dose finding studies, together with the associated MED (assuming an improvement of 200 units over the placebo response as the clinical relevance threshold). As seen from Figure 2-1, the MED depends quite strongly on the true, underlying dose response profile and can vary between 50 (for the *emax1* model) and 350 (for the *linear* model).

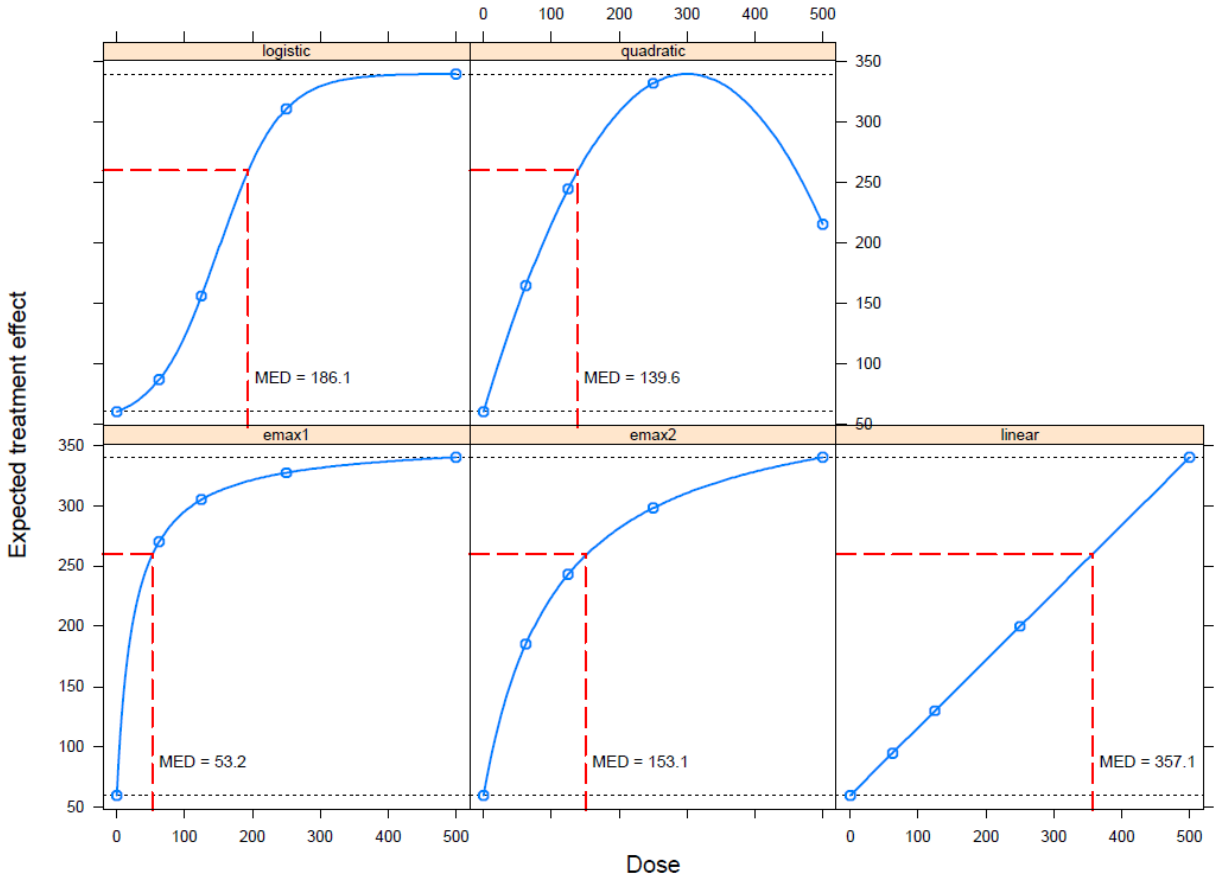

Figure 2-1 **Example dose response profiles often seen in clinical dose finding studies. Open dots indicate the expected responses at selected dose levels. The minimum effective dose (MED) is defined as the smallest dose achieving a pre-specified clinical treatment.**

In practice there is not much knowledge about the true underlying dose response profile at the time the study is being designed, so the initial dose specification is tentative at best. Thus, there is a need to develop efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty.

2.2 Objectives of Phase II dose finding studies

An indication of the importance of properly conducted (and informative) dose response studies is the early publication of the ICH E4 guideline (ICH, 1994), which is the primary source of regulatory guidance in this area. The guideline gets very specific already in the introduction when it motivates the importance of dose response information:

> *Historically, drugs have often been initially marketed at what were later recognized as excessive doses ... This situation has been improved by attempts to find the smallest dose with a discernible useful effect or a maximum dose beyond which no further beneficial effects is seen...*

It becomes transparent from this quote, and the remainder of the ICH E4 guideline, that regulatory agencies recognize the need to obtain appropriate dose response information as a

critical part of clinical drug development. But even if it is generally agreed that understanding the relationships among administered dose, drug-concentration in blood, and clinical response is important, setting the objectives for an actual trial may be subject to much debate. For example, Ruberg (1995a, b) and subsequently Bretz et al. (2008) considered the following questions to be relevant in the context of dose finding:

1. *Is there any evidence of a drug effect?*

   The detection of a dose response signal is often related to the determination of proof-of-concept (PoC) in a development program. This is a critical decision point, since a positive PoC coupled with a subsequent commitment to go into full development leads to substantial investment of resources.

2. *What doses are (relevantly) different from control?*

   This question is closely connected to the estimation of a minimum effective dose, that is, "the smallest dose with a discernible useful effect" (ICH, 1994). If confirmatory pairwise comparisons with a control are of main interest (such as in Phase III trials), multiple comparison procedures may be appropriate to answer this question.

3. *What is the dose response relationship?*

   This question is broader than the previous one in the sense that it asks for a complete functional description of the dose response relationship. If this is of main interest, modeling approaches may be appropriate to take full advantage of the observed data. Understanding the dose response relationship is particularly important when it is necessary to pick a lower dose because of safety concerns observed in an actual study.

4. *What is the optimal dose?*

   Although very natural, this question is likely to be the most difficult to answer. In practice this question may not even be well defined in the sense that different stakeholders may have a different understanding of what "optimal dose" means. In all circumstance, any answer to this question will be a trade-off between efficacy considerations, tolerability and safety concerns and regimen convenience.

2.3 Overview of statistical analysis methods

The analysis of dose finding studies can be classified into two major strategies: multiple comparison procedures (Bretz et al., 2010) and modeling techniques (Pinheiro et al., 2006a).

Multiple comparison procedures regard the dose as a qualitative factor and make few, relatively weak assumptions about the underlying dose response profile. Multiple comparison procedures can be used for detecting an overall dose related signal, as well as for estimating target doses of interest. Stepwise testing strategies can be applied which preserve the overall Type I error rate at a pre-specified level α. Such procedures are relatively robust to the underlying dose response profile, but they are not designed for extrapolation of information beyond the observed dose levels. Inference is thus confined to the selection of the target dose among the dose levels under investigation. Multiple comparison procedures are often used to address Questions 1 and 2 from Section 2.2.

Modeling approaches make better use of the available information as they interpolate information across doses instead of treating every dose separately. They hence consider dose as the continuous variable it truly is. The drawback of modeling approaches is their dependence on the assumed model: Any inference (e.g. target dose estimation) depends on the employed dose response model and can be highly sensitive to its choice. Because the dose response model and its parameters are unknown prior to a clinical study, one is faced with model uncertainty, a problem which is often underestimated. A common approach is to fit several dose response models once the data have been observed and select the best fitting model. However, such a naïve approach does not account for model uncertainty and can lead to undesirable effects due to data dredging, such as overfitting, biased treatment effect estimates and over-optimistic analysis results; see Chatfield (1995), Draper (1995), and Hoeting et al. (1999).

Hybrid dose finding methods that combine principles of multiple comparisons with modeling techniques have recently been investigated to overcome some of the shortcomings of applying either approach alone. An early reference is Tukey et al. (1984), who recognized that the power of standard hypotheses tests to detect a dose response signal depends critically on the unknown dose response relationship and developed trend tests to take this uncertainty into account. They proposed to simultaneously use several trend tests and subsequently to adjust the resulting p-values for multiplicity. Bretz et al. (2005) proposed an extension of this methodology, denoted MCP-Mod, which provides the flexibility of modeling for dose estimation, while preserving the robustness to model misspecification associated with multiple comparison procedures. The principles underlying the MCP-Mod approach are outlined in Section 2.4. A detailed technical description with several extensions is given in Section 3.

Some authors have investigated semi-parametric or non-parametric approaches to alleviate the model dependency problem and enhance the robustness of the dose response estimation; see Kelly and Rice (1990); Mukhopadhyay (2000); Dette et al. (2005); Bornkamp and Ickstadt (2009); Yuan and Yin (2011) among many others. These methods allow descriptions of dose response relationships that do not depend on fully specified parametric models. However, their applicability in dose response studies is limited because some of these methods require observations on a rather dense set of different dose levels, which are rarely available in practice. Due to logistic and ethical reasons, the number of different dose levels is often in the order of 5 (Bornkamp et al., 2007) and therefore nonparametric methods may not yield reliable results in the context of clinical dose finding studies.

2.4 Principles of the MCP-Mod approach

The MCP-Mod approach provides a unified strategy for addressing the key questions in dose finding trials, as outlined in Section 2.2. More specifically, it provides an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies acknowledging model uncertainty through the following steps: (i) testing for the presence of a dose response signal, (ii) selecting the best dose response model for the observed data out of a pre-specified set of candidate models, and (iii) estimating target doses of interest (e.g., the minimum effect dose, MED) via modeling.

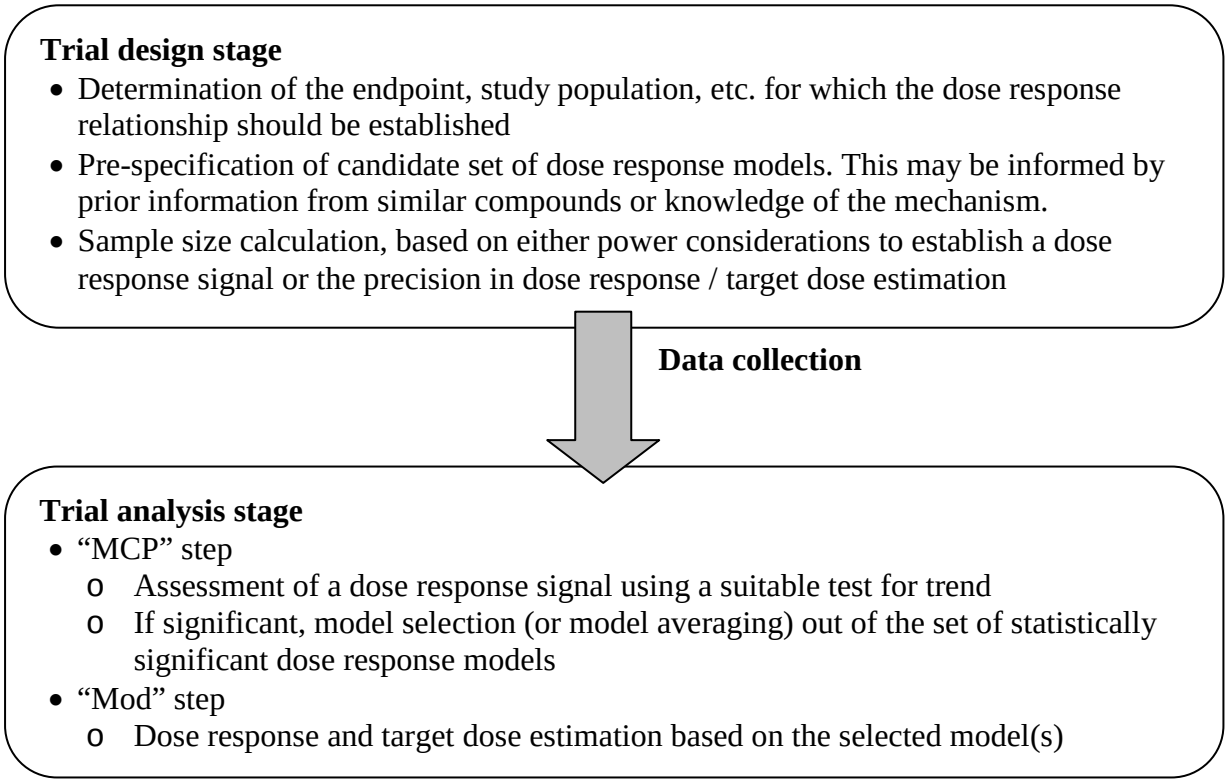

Figure 2-2 High-level overview of the MCP-Mod approach

Figure 2-2 gives a high-level overview of the MCP-Mod approach, showing the steps needed at the trial design stage and the trial analysis stage. At the trial design stage the clinical team needs to decide on the core aspects of the trial design, as in any other trial. Specific to the MCP-Mod approach is the need to pre-specify a candidate set of plausible dose response models, in discussion with the clinical team, based on available pharmacologic information, dose response information from similar compounds, data from earlier trials etc.; see Table 6-1 in the Appendix for a list of commonly used dose response models. This candidate set then gives rise to a set of optimal contrasts used to test for the presence of a dose response signal (see Section 3 for technical details). In case of large model uncertainty, this candidate set should be chosen to cover a large and diverse set of plausible dose response shapes. Sample size calculations at the trial design stage could be based either on power considerations (e.g. to achieve a pre-specified probability of establishing a true dose response signal) or on a pre-specified precision to estimate the dose response relationship or a target dose of interest (e.g. the expected confidence interval width for the MED estimate).

The trial analysis stage consists of two main steps: The MCP and the Mod step. The MCP step focuses on establishing evidence for a drug effect across the doses, i.e. detecting a statistically significant dose response signal for the clinical endpoint and patient population investigated in the study. This step will typically be performed using an efficient test for trend, adjusting for the fact that multiple candidate dose response models are being considered. If a statistically significant dose response signal has been established, one proceeds with determining a reference set of significant dose response models by discarding the non-significant models

from the initial candidate set. Out of this reference set, a best model is selected for dose estimation in the last stage of the procedure. The selected dose response model is then employed to estimate target doses and possibly incorporating information on clinically relevant effects. The precision of the estimated doses can be assessed using, for example, bootstrap methods.

In contrast to a direct application of modeling dose estimation, the MCP step accounts for possible model mis-specification and includes the associated statistical uncertainty through a hypothesis testing context. Note that different model selection criteria may lead to different dose estimates due to different sources of information and/or decision goals: Information criteria, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), are statistical decision rules taking only the data from the study into account; Bayesian decision rules may additionally include information external to the study, though still based on statistical reasons; non-statistical selection rules may also be applied and can be based on updated clinical knowledge, economic reasons, etc.

The MCP-Mod approach displayed in Figure 2-2 is very flexible and can be tailored to the actual needs of a given Phase II study. For example, if substantial prior knowledge about the dose response model is available, the candidate set could consist of only one single dose response model (e.g. Emax). In such cases, the main goal might be to establish a dose response signal or to estimate the rough shape of the dose response relationship. At the trial analysis stage, there are several possibilities of performing an efficient trend test. The multiple contrast test approach described in Section 3 has several advantages, although other methods are also available, such as employing likelihood ratio tests (Dette et al., 2013). In some applications, the number of doses might not be sufficient to support the Mod step. Also, instead of a selecting a single dose response model and proceed with the Mod step, one could use one of several model averaging techniques (Buckland et al., 1997). In Section 3 we provide the technical details of the original proposal by Bretz et al. (2005), which focused on a single normally distributed endpoint, and further describe several variants thereof.

2.5 In-scope and out-of-scope of this request

In its currently available version, the MCP-Mod methodology is best used in trials satisfying certain characteristics:

- *Drug development stage*: Phase II dose finding studies to support dose selection for Phase III
- *Response*: Univariate (efficacy or safety/tolerability) variable. For efficacy, the response variable is ideally predictive to the clinical Phase III efficacy outcome. Could be a binary, count, continuous or time-to-event endpoint. Observations could be cross-sectional (i.e. from a single time point) or longitudinal.
- *Dose*: Typically, the dose levels utilized in the actual trial are used for the design and analysis. However, more broadly “dose” could be any univariate, continuous, quantitative measurement, as long as an ordering of the measurements is possible and the differences between measurements are interpretable. For example, sometimes it is possible to convert b.i.d. and o.d. regimen to a common univariate scale.

- *Number of doses*: For the Mod step, a minimum of four distinct doses (including placebo) is required, ideally distributed over the effective range. For the MCP step (e.g. for dose response signal testing or identifying the type of plausible dose response shapes), at least three distinct doses (including placebo) are needed.

The framework covered in this request includes parallel group and more complex (crossover, etc.) designs. Applications to studies without a placebo group and/or with an active control group are possible. One or multiple interim analyses (e.g. for futility stopping or response-adaptive dose allocations) are possible (and often advisable).

Examples of trial designs and modeling approaches which are out of scope for this request or where only limited experience is available include, among others:

- Predictions from a surrogate / biomarker or short term readout to a clinical Phase III endpoint.
- Exposure-response analyses or PK-PD models are possible (if appropriate models are available) but not the purpose of this request per se.
- Titration designs and dose escalation studies (e.g. to estimate the maximum tolerable doses using continual reassessment methods).
- Regimen finding for long acting biologics where there is no steady state.
- Application of MCP-Mod in confirmatory studies.
- Multivariate problems, such as the joint modeling of efficacy and toxicity, the presence of two primary endpoints, or drug combination trials.

Although, as of yet, there is limited experience with the use of MCP-Mod in pediatric trials at Novartis (or other companies we are aware of), the methodology is potentially as useful in this setting as in adult dose finding. In fact, the more stringent ethical and resource constraints typically present in pediatric development programs make the use of information-efficient methods like MCP-Mod even more appealing. Typically only few patients and few data are available for pediatric trials. To increase the efficiency of pediatric dose-finding studies different alternatives could be envisaged like including age (or weight) as a covariate in the model, adaptive dose-finding studies with interim analyses, and the use of longitudinal data and dose-exposure response studies. More discussions and, possibly, methodological research are needed to further promote the use of MCP-Mod in pediatric dose finding programs and its inclusion in Pediatric Investigational Plans (PIPs).

3 Methodology and results

3.1 Methods

As a concrete illustration of the general MCP-Mod approach described in Section 2.4, we provide in Section 3.1.1 a short, technical description of the original MCP-Mod procedure (Bretz et al., 2005) for a single, normally distributed efficacy endpoint, and describe its implementation in the `DoseFinding` package in R, which is freely available from www.r-project.org. In Section 3.1.2 we describe recently developed variants of MCP-Mod that

include adaptive dose finding techniques and dose finding for general parametric models (e.g. for time-to-event endpoints and including longitudinal data modeling).

3.1.1 Original MCP-Mod procedure

The original MCP-Mod procedure for a single, normally distributed efficacy endpoint is implemented in five steps; see also Figure 3-1, which complements the high-level overview from Figure 2-2 for this concrete case.

1. Identify several candidate parametric models, which are likely to represent the underlying dose response shape.
2. Derive optimum contrast coefficients, such that the marginal power to detect a specific dose response shape associated with the respective candidate model is maximized.
3. Evaluate the significance of the individual models in terms of a multiple contrast test based on the previously derived optimal contrast coefficients.
4. Select a candidate model associated with the most significant contrast test, or other model selection criteria such as AIC or BIC (provided statistical significance has been shown in the previous step). Alternatively, multiple significant models can be selected if model averaging is preferred.
5. Use the selected model(s) to produce inferences on adequate doses, employing a model-based approach.

The results from this analysis may then form the quantitative basis for selecting the dose for the Phase III program. In practice, selection of one or more doses to advance into Phase III clinical trials is one of the most challenging decisions during drug development and will typically involve both qualitative and quantitative considerations.

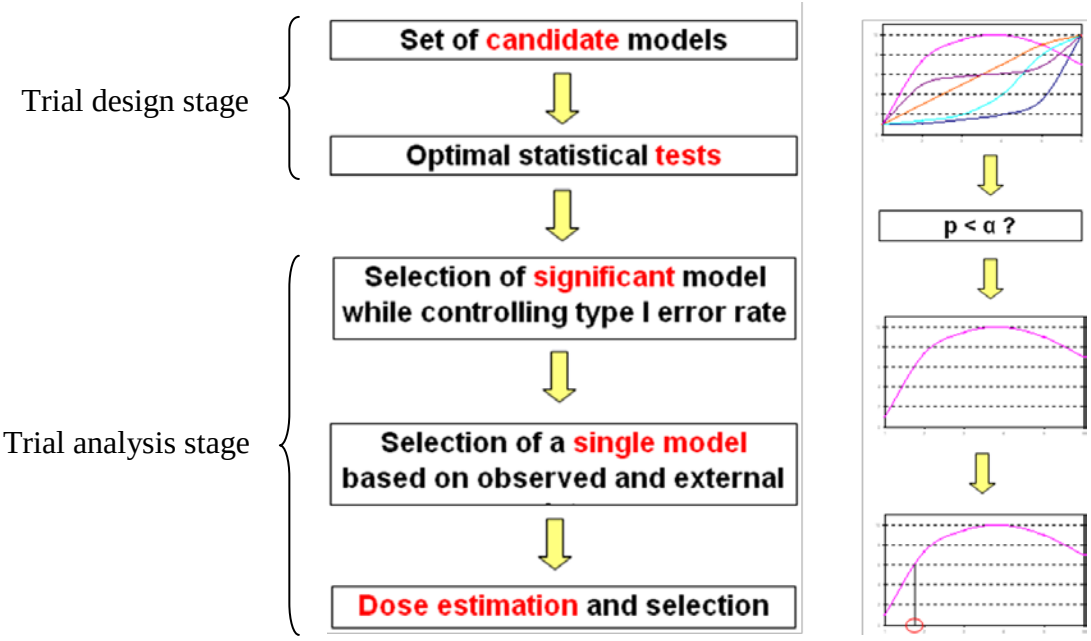

Figure 3-1 **High-level overview of the original MCP-Mod procedure from Bretz et al. (2005) for a single, normal distributed efficacy endpoint.**

In the following we describe the individual steps of the MCP-Mod approach in more detail. To this end, assume a primary response variable y for a given set of parallel groups of patients corresponding to active doses d_2, d_3, ..., d_k plus placebo d_1, for a total of k arms. For the purpose of dose response signal testing and estimating target doses, we consider a one-way layout for the model specification

$$y_{ij} = \mu_i + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0, \sigma^2),\ \ j = 1, \ldots, n_i,\ i = 1, \ldots, k,$$

where the mean response at dose d_i can be represented as $\mu_i = \mu(d_i, \boldsymbol{\theta})$ for some dose response model $\mu(.)$ parameterized by a vector of parameters $\boldsymbol{\theta}$, n_i denotes the number of patients allocated to dose d_i, and ε_{ij} is the error term for patient j within dose group i (assuming independent residuals). We note that most dose response models used in practice can be written as $\mu(d, \boldsymbol{\theta}) = \theta_0 + \theta_1 f^0(d, \boldsymbol{\theta}_2)$, and $f^0(d, \boldsymbol{\theta}_2)$ is a, typically a nonlinear, transformation of the dose levels depending on $\boldsymbol{\theta}_2$. The parameters θ_0 and θ_1 determine location and scaling of the function f. An example for f^0 is $f^0(d, \boldsymbol{\theta}_2) = d/(d + \boldsymbol{\theta}_2)$, resulting in the (hyperbolic) Emax model and where $\boldsymbol{\theta}_2$ is the ED_{50} parameter of the Emax model. More dose response models are given in Table 6-1 in the Appendix.

The first step of MCP-Mod is to identify a set of M parameterized candidate models together with prior parameter values for their standardized versions. When used with the doses planned for the trial, $d_1, \ldots, d_k$, these candidate models produce mean response vectors $\boldsymbol{\mu}_m = (\mu_{m1}, \ldots, \mu_{mk})$, where $\mu_{mi} = \mu_m(d_i, \boldsymbol{\theta}_{2m})$ with a specified parameter $\boldsymbol{\theta}_{2m}$ and μ_m is the m-th dose response function, m = 1, …, M.

The second step is to test each of the dose response shapes in the candidate set using a single contrast test, with coefficients chosen to maximize the power of the test when the true underlying mean response equals $\boldsymbol{\mu}_m$. This approach formulates the test for a dose response

trend in terms of a linear contrast hypothesis $\mathbf{c}'\boldsymbol{\mu}$, where $\mathbf{c}$ is a contrast vector subject to $\Sigma_i c_i = 0$ and $\boldsymbol{\mu}$ is the true mean at the dose levels. For a given vector $\boldsymbol{\mu}$, optimum contrast coefficients for model testing are proportional to $n_i(\mu_i - \mu^*)$, i = 1, …, k, which after normalization lead to the unique solution (up to the sign) $\mathbf{c}/\|\mathbf{c}\|$, where $\mu^* = \boldsymbol{\mu}'\mathbf{1}/k$. Each of M possible dose response shapes now gets represented by one contrast $\mathbf{c}_m = (c_{m1}, \ldots, c_{mk})'$. The contrast coefficients for the m-th shape are specified at the trial design stage, such that they maximize the power to detect the m-th expected dose response shape. For example, a linear contrast test for equally spaced doses and balanced patient allocation is defined such that the difference of any two adjacent contrast coefficients is a constant. Assuming that the linear model has been included in the candidate model set, the linear contrast test is then a powerful test to detect the linear trend. Similarly, any dose response relationship characterized through $\boldsymbol{\mu}_m$ can be tested equally powerfully by selecting an appropriate contrast test, whose coefficients are defined in dependence of the assumed $\boldsymbol{\mu}_m$. Note that contrast tests are shift and scale invariant, and thus it is sufficient to work with the standardized modeling functions f.

The third step is to test for an overall dose response signal. The single contrast tests are defined as

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

where $S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2/(N-k)$ denotes the mean squared error and $N = \Sigma_i n_i$ the total sample size. Under the null hypothesis H: $\mathbf{c}'\boldsymbol{\mu} = 0$ of no dose response effect, i.e. $\mu_1 = \ldots = \mu_k$, and under the distributional assumptions from above, $T_1, \ldots, T_M$ jointly follow a central multivariate t distribution with N – k degrees of freedom and the correlation matrix only depends on the sample sizes and the correlations between the different contrasts. Note, that for the data analysis step the critical value, contrast weights, etc. cannot be taken over from the design stage due to drop-outs, covariate effects, etc. In practice these computations have therefore to be re-done at this point.

The final test statistic T_{max} is based on the maximum contrast test, i.e., $T_{max} = \max_m T_m$. A dose response signal is verified if this maximum statistic T_{max}, and thus at least one single contrast test, is statistically significant while controlling the familywise error rate (FWER) at pre-specified level α. Let $q_{1-\alpha}$ denote the multiplicity adjusted critical value. A dose response signal is then established if $T_{max} \geq q_{1-\alpha}$. In fact, any dose response model with a test statistic larger than $q_{1-\alpha}$ can be declared statistically significant at level α. These models then form a reference set $\{M_1, \ldots, M_L\}$ of L significant models, provided $L \geq 1$. Every single contrast test thus translates into a decision procedure, whether a selected dose response curve is significant given the observed data, while controlling the FWER at pre-specified level α. If no candidate model is statistically significant, the MCP-Mod procedure stops indicating that a dose response signal cannot be established from the observed data. But such a result does not necessarily mean that the compound has no effect at all. In general we will assess that there was not enough evidence to get statistical significance. Possible reasons could include small sample sizes or high variance. In addition, the initial candidate set might have been poorly chosen, such that the candidate models do not fit the true curve (perhaps an umbrella-shaped

curve) – although in such cases the compound has a pronounced effect and may be far from ineffective, as suggested by the insignificant p-value.

The fourth step is to select one model out of the reference set of L significant models, provided significance has been shown in the previous step. The selected model could be associated with the most significant contrast test, or other model selection criteria such as AIC or BIC. Alternatively, multiple significant models can be selected if model averaging is preferred (Buckland et al., 1997). Note that in contrast to a direct application of a model based approach (which can lead to biased estimates and over-optimistic analysis results; see Section 2.3), the preliminary steps of the MCP-Mod procedure address issues of possible model misspecifications and include the associated statistical uncertainty in a rigorous hypothesis testing framework.

The fifth and final step consists of fitting the selected model to the data and estimating adequately the target dose(s) of interest using standard inverse nonlinear regression techniques.

3.1.1.1 DoseFinding package in R

We illustrate the basic principles of MCP-Mod using a simple numerical example with a homoscedastic normally distributed endpoint. To this end, we use the `DoseFinding` package in R, which provides an implementation of the MCP-Mod approach. A compiled version and the source code of the package are freely available from the Comprehensive R Archive Network (CRAN) servers; see http://cran.r-project.org/package=DoseFinding. This ensures reproducibility and transparency of the underlying calculations. The `DoseFinding` extends the earlier `MCPMod` package described in Bornkamp at al. (2009). The package includes functions for the design and analysis of dose finding trials. All code and analyses were produced with version 0.9-6 of the `DoseFinding` package.

Currently, the main functions to design a dose finding study include:

- `powMCT:` Calculates the power of a multiple contrast test under specified dose response shapes.
- `sampSizeMCT:` Calculates the sample size necessary to achieve a specific power for a given multiple contrast test under pre-specified dose response shapes.
- `optDesign:` Calculates optimal designs (allocation of doses and patient allocation weights) for dose response estimation (D-optimality) and target dose estimation (TD-optimality).

The main analysis functions in the package are:

- `MCTtest`: Performs multiple contrast tests for dose response signal detection.
- `fitMod`: Fits nonlinear dose response models and provides the estimated dose response curve.
- `MCPMod`: Wrapper function that calls `MCTtest`, then fits dose response models using the `fitMod` function and selects one (or more) suitable models according to a

specified model selection criterion. Finally, it provides estimates of the dose response curve and target doses of interest.

The `DoseFinding` package hence provides all the tools necessary to design and analyze a dose finding study using MCP-Mod. It is written in a modular form that can easily be extended to cover non-standard situations, as illustrated with the example in Section 3.1.2.3.

3.1.1.2 Example

We use the `biom` dataset from the `DoseFinding` package to illustrate the MCP-Mod methodology. The data set contains 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 per group. The response variable was assumed to be normally distributed and larger values indicate a better outcome.

The first step of MCP-Mod is to identify the candidate model set, which can be done with the `Mods` function. In practice, identifying the candidate model set is the result of a highly interactive and iterative clinical team discussion, using all available information and plausibility arguments. Here, we select the candidate set to include two concave shapes being the linear-in-log and the quadratic shape (with δ = -0.83, leading to an umbrella shape with the maximum treatment effect occurring at dose 0.6), a convex shape using an exponential (δ = 0.4), and the linear shape; see Table 6-1 in the Appendix for details on the model parameterizations. The shapes can be specified and visualized with `DoseFinding` using the following code:

```
> ## load DoseFinding package
> library(DoseFinding)
> doses <- c(0,0.05,0.2,0.6,1) ## doses in the study
> ## define candidate set
> candMod <- Mods(linlog=NULL, linear=NULL, quadratic=-0.83, exponential=0.4, doses=doses)
> ## plot candidate set
> plot(candMod, placEff = 0, maxEff = 1)
```

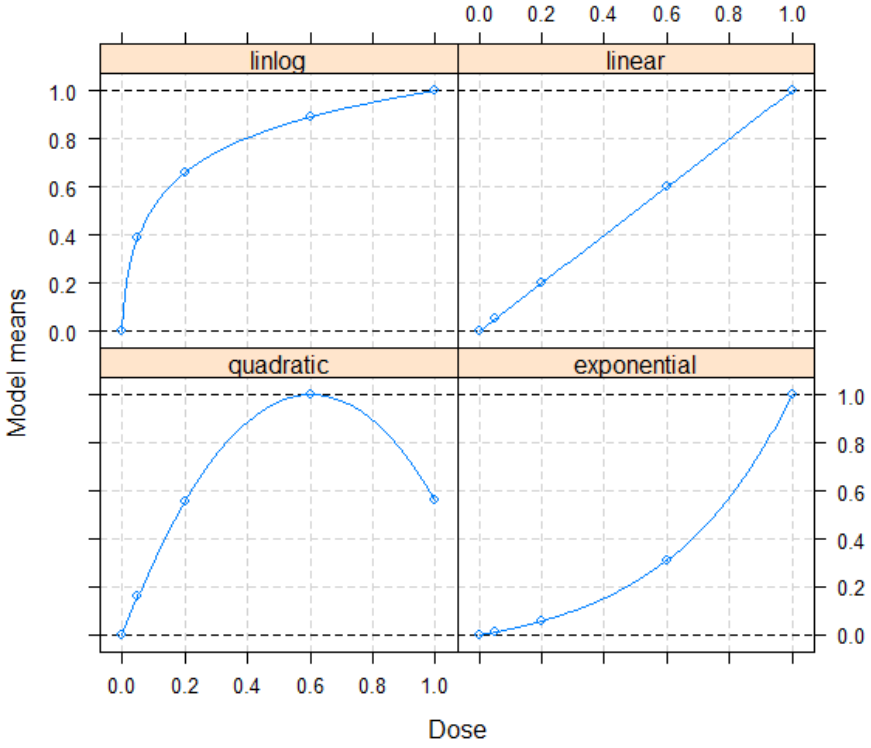

Figure 3-2 Display of candidate candidate dose-response shapes

After the study is completed and data have been obtained, one can use the `MCP-Mod` function to analyze the data. The function needs as input the dose and response values, the candidate models and the clinical relevance threshold Δ to perform target dose estimation (in this example we assume Δ = 0.4). To run the MCP-Mod analysis, the following code can be used:

```
> data(biom) ## load data
> ## fit MCPMod
> MMfit <- MCPMod(dose, resp, models = candMod, data=biom, Delta = 0.4)
> MMfit
MCPMod


Multiple Contrast Test:
            t-Stat   adj-p
linlog       3.411 0.00104
quadratic    3.202 0.00206
linear       2.972 0.00442
exponential  2.418 0.02010

Estimated Dose Response Models:
linlog model
   e0 delta
0.975 0.146
```

```
linear model
   e0 delta
0.492 0.559

quadratic model
    e0     b1     b2
 0.390  1.768 -1.232

exponential model
   e0    e1 delta
0.511 0.833 2.000

Selected model (AIC): linlog

Estimated TD, Delta=0.4
     linlog      linear   quadratic exponential
     0.1455      0.7161      0.2813      0.7843
```

One can observe that the linear-in-log shape has the largest contrast test statistic but also all other dose response shapes are significant. Hence, all dose response models proceeded to the model estimation step, with the estimated model parameters been shown in the output. The linear-in-log model is ultimately selected when applying AIC model selection to determine a single dose response model out of the selected model set. The estimated target doses, giving an improvement of 0.4 over placebo under the different models are shown in the last line of the output. More information on the fitted models can be obtained by calling `summary(MMfit)` or by directly observing the content of the `MMfit` object. A graphical depiction of the fitted dose response model can be obtained via:

```
## plot resulting model-fit
plot(MMfit$mods$linlog, CI=TRUE, plotData = "meansCI", lwd=2)
```

This concludes the MCP-Mod analysis. In practice, these study results, together with further dose response analyses on other efficacy and/or safety endpoints, would provide the quantitative basis for (i) the decision to advance into Phase III and (ii) the dose(s) being selected in future studies.

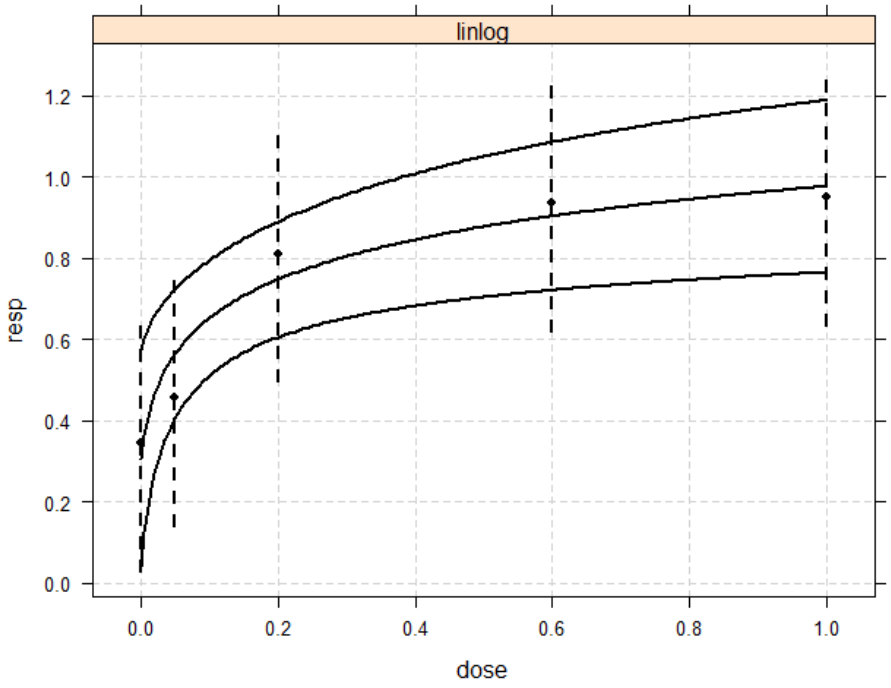

Figure 3-3 **Linear in log model fitted to the biom dose-response data with pointwise 95% confidence intervals for the fitted curve and the corresponding ANOVA based pointwise 95% confidence intervals**

3.1.2 Variants

The original MCP-Mod procedure from Bretz et al. (2005) was motivated by the work of Tukey et al. (1984), who proposed to simultaneously use several trend tests based on different functional dose response descriptions and to subsequently adjust the resulting *p*-values for multiplicity. Since then, the methodology has been subject of several investigations. Pinheiro et al. (2006b) discussed practical considerations regarding the implementation of this methodology, including sample size calculations for the MCP part. Neal (2006) and Wakana et al. (2007) investigated extensions to Bayesian methods for estimating or selecting the dose response curve from a sparse dose design. Dette et al. (2008) constructed optimal designs for MED estimation to determine the optimum location of doses and allocation of patients to the individual dose levels, taking model uncertainty into account. Klingenberg (2009) applied the MCP-Mod approach to proof-of-concept studies with binary responses. Benda (2010) proposed a time-dependent dose finding approach with repeated binary data. Akacha and Benda (2010) investigated the impact of dropouts on the analysis with recurrent event data. Tao (2010) applied the MCP-Mod procedure in joint modeling of efficacy and safety endpoints in Phase II studies. Wouters (2012) extended the MCP-Mod approach to investigate longitudinal toxicological data. Dette et al. (2013) investigated the use of likelihood ratio tests instead of contrast tests. Miller (2010), Tanaka (2011) and Tanaka and Sampson (2013) extended the MCP-Mod approach to flexible two-stage designs that include the possibility to perform adaptations (e.g. adding or dropping doses) under a strict FWER control. In the following we describe two variants of the MCP-Mod procedure in more detail. First, we describe its use in response-adaptive dose finding studies. Second, we describe its use for

general parametric models (e.g. for time-to-event endpoints and including longitudinal data modeling).

3.1.2.1 Response-adaptive dose finding

Dragalin et al. (2010) and Bornkamp et al. (2011) investigated the MCP-Mod procedure in response-adaptive designs, addressing two major challenges in dose-finding studies: uncertainty about the dose response models and large variability in parameter estimates. To allocate new cohorts of patients in an ongoing study, optimal designs are used that are robust under model uncertainty. In addition, a Bayesian shrinkage approach is used to stabilize the parameter estimates over the successive interim analyses used in the adaptations. This approach allows calculating updated parameter estimates and model probabilities that can then be used to calculate the optimal design for subsequent cohorts. The resulting designs are hence robust with respect to model misspecification and additionally can efficiently adapt to the information accrued in an ongoing study. Bornkamp et al. (2011) focused on adaptive designs for estimating the minimum effective dose, although alternative optimality criteria or mixtures thereof could be used, enabling the design to address multiple objectives; see Dragalin et al. (2010). An alternative approach to use MCP-Mod in adaptive dose finding is to use decision rules that are fine-tuned to the specific trial considerations. For example, Mercier et al. (2013) described a real case study of a two-stage design. For decision making at interim different hypothetical scenarios of the observed dose response shape were set up, and for each scenario a corresponding design for the second stage was pre-specified. At the interim analysis one then observes to which scenario the observed data correlate best, and the corresponding design is then selected for the second stage of the trial. An example of such a rule could be to include a lower dose at interim if all active doses at the interim are on the plateau of the dose response curve, which was the case in the actual study described in Mercier et al. (2013).

3.1.2.2 MCP-Mod for general parametric models

In practice the modeling situation and study design are more complex than described in Section 3.1.1. For example, the response variable might be a count, binary or time-to-event variable instead of being normally distributed. In addition, the final analysis has usually to be adjusted for relevant covariates (e.g. region, age, …), patients measurements are often recorded over time (necessitating the use of longitudinal models), and patients might receive more than one treatment (such as in cross-over or incomplete block designs). The core ideas of MCP-Mod remain applicable in these situations, as sketched in the following and laid out in Pinheiro et al. (2013) in more detail.

Let $\mathbf{y}$ denote the response vector of an experimental unit in the trial (e.g., a patient) which has been assigned a dose d (the formulation can easily be extended to the case of multiple doses). We assume that the residual distribution function is given by $\mathbf{y} \sim F(\mathbf{z}, \boldsymbol{\eta}, \mu(d))$ where $\mu(d)$ denotes the dose response parameter, $\boldsymbol{\eta}$ the nuisance parameters, and $\mathbf{z}$ possible covariates. The key features of MCP-Mod can then be formulated with respect to $\mu(d)$, including:

- accounting for uncertainty in the dose response model via a set of candidate dose response models,

- testing for a dose response signal via optimal contrasts based on plausible dose response shapes,
- model selection or model averaging to combine different models, and
- dose response estimation and dose selection using nonlinear regression.

Because all dose response information is assumed to be represented by μ(d), the interpretability of this parameter is critical for communicating with clinical teams, choosing candidate dose response shapes, specifying clinically relevant effects, etc. As an example consider the Weibull distribution: It is typically parameterized by a scale parameter λ and shape parameter α, neither of which is easily interpretable. For the purpose of interpretability, the model could be re-parameterized in terms of the median time to event $\mu = \log(2)^{1/\alpha}/\lambda$ and α, and then use μ as an interpretable dose response parameter.

Let $\hat{\mu}$ denote the vector of estimated dose response parameters under an analysis-of-variance (ANOVA) parameterization, obtained using the appropriate estimation method for the general parametric model above via maximum likelihood (ML), generalized estimating equations, partial likelihood, etc. The key assumption needed for this general version of MCP-Mod to work is that one can extract estimates of the dose response of form $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{S})$, where **S** denotes the variance-covariance of $\hat{\boldsymbol{\mu}}$. This assumption can be shown to hold for most parametric estimation problems, such as, generalized linear models, parametric time-to-event models, mixed-effects models, etc. The MCP step consists of specifying a set of candidate models for the dose response relationship μ(d). Similar to the original MCP-Mod approach, each of the candidate model shape determines an optimal contrast for a trend test to evaluate the associated dose response model signal. The optimal contrasts are applied to the previously described ANOVA-type estimates $\hat{\mu}$, with the associated asymptotic distribution used for implementing the corresponding tests (i.e., critical values and p-values). It can be shown that the (optimal) contrast for testing the hypothesis that $\mathbf{c}'\boldsymbol{\mu} = \mathbf{0}$ with maximal power for a single given candidate model shape $\boldsymbol{\mu}_m$ is given by $\mathbf{S}^{-1}(\boldsymbol{\mu}_m - (\boldsymbol{\mu}_m'\mathbf{S}^{-1}\mathbf{1})/(\mathbf{1}'\mathbf{S}^{-1}\mathbf{1}))$, which generalizes the formula given in Section 3.1.1. Once a dose response signal is established, one proceeds to the Mod step, fitting the dose response profile and estimating target doses based on the models identified in the MCP step. There are many ways to fit the dose response models to the observed data, including approaches based on maximizing the likelihood or the restricted likelihood. Pinheiro et al. (2013) suggested an alternative two-stage approach to dose response model fitting based on generalized least squares, which has some computational advantages. Although this approach relies on asymptotic results, it has the appeal of being a general purpose application, as it depends only on $\hat{\mu}$ and $\hat{S}$, and is implemented in the `DoseFinding` package in R.

3.1.2.3 Example

This example is taken from Pinheiro et al. (2013) and illustrates some of the generality of MCP-Mod described in Section 3.1.2.2 as well as its implementation in R using the `nlme` and `DoseFinding` packages. This example refers to a Phase II clinical study of a new drug for a neurodegenerative disease. The state of the disease is measured through a functional scale, with smaller values corresponding to more severe neurodeterioration. The goal of the drug is to reduce the rate of disease progression, which is measured by the linear slope of the

functional scale over time. We first describe the statistical model and then illustrate the MCP-Mod procedure with a simulated dataset.

The functional scale response is assumed to be normally distributed and, based on historical data, it is believed that the longitudinal progression of the functional scale over the one year of follow up can be modeled by a simple linear trend. An alternative analysis would be to use the change from baseline after one year as an endpoint and use the methods described in Section 3.1.1, although such a cross-sectional analysis would not use all available data.

We consider a mixed-effects model representation for the functional scale measurement y_{ij} on patient i at time t_{ij}:

$$y_{ij} = (\beta_0 + b_{0i}) + (\mu(d_i) + b_{1i})t_{ij} + \varepsilon_{ij},\ (b_{0i},b_{1i})' \sim N(\mathbf{0}, \mathbf{\Lambda}) \text{ and } \varepsilon_{ij} \sim N(0, \sigma^2).$$

That is, every patient suffers over time (t) from a disease progression on the functional scale according to a linear regression with patient specific intercept $\beta_0 + b_{0i}$ and slope $\mu(d_i) + b_{1i}$, where the slope depends on the assumed dose via $\mu(d)$. If $\mu(d)$ is represented by a linear function of dose d, this is a linear mixed-effects (LME) model, else it becomes a nonlinear mixed-effects (NLME) model. The dose response parameter in this case is the linear slope of disease progression $\mu(d)$ on which the MCP-Mod procedure will be applied.

The research interest in this study focuses on the treatment effect on the linear progression slope. Without going into the detailed rationale for the chosen doses and sample size, we assume that the trial design includes placebo and four doses, 1, 3, 10, and 30 mg, with balanced allocation of 50 patients per arm. Patients are followed up for one year, with measurements of the functional scale being taken at baseline and every three months thereafter. The study goals are to (i) test the dose response signal, (ii) estimate the dose response and (iii) select a dose to be brought into the confirmatory stage of the development program.

At the planning stage of the trial, the following assumptions were agreed with the clinical team for the purpose of the design:

- Natural disease progression slope is –5 points per year.
- Placebo effect is 0 (i.e., no change in natural progression).
- Maximum improvement over placebo within dose range is a 2 points increase in slope over placebo.
- Target (clinically meaningful) effect is 1.4 points increase in slope over placebo.

Plausible values for the variance-covariance parameters were obtained from historical data as follows: $\text{var}(b_{0i}) = 100$, $\text{var}(b_{1i}) = 9$, $\text{corr}(b_{0i},b_{1i}) = -0.5$, and $\text{var}(\varepsilon_{ij}) = 9$. The ANOVA-type estimate $\widehat{\mu}$ of $\boldsymbol{\mu} = (\mu_{0mg}, \mu_{1mg}, \mu_{3mg}, \mu_{10mg}, \mu_{30mg})'$ will hence consist of the linear progression slopes for the 4 treatment groups. For interpretability it might be simpler to think of it as change from baseline (since for t = 1 these two quantities are identical).

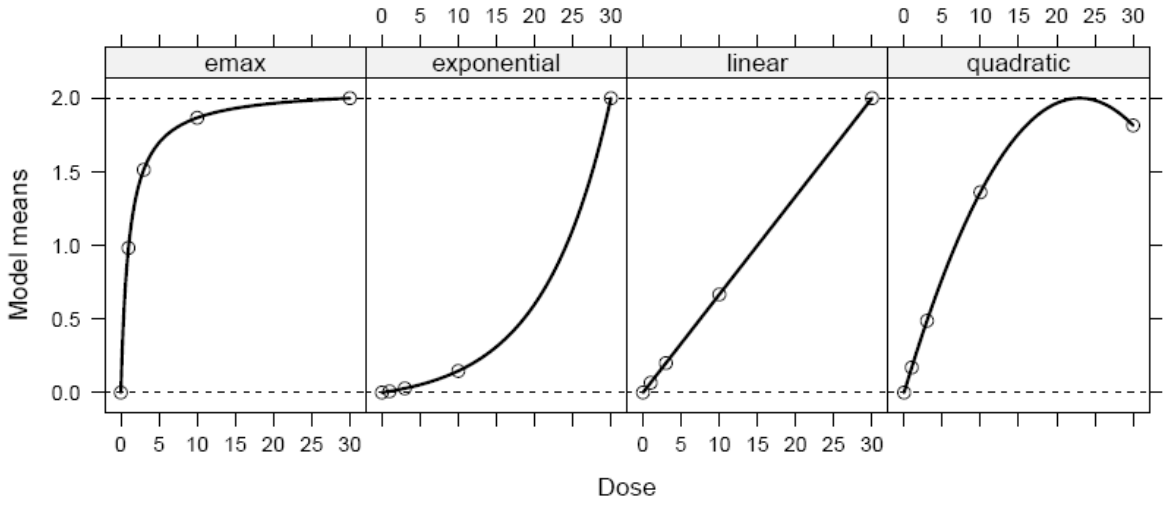

Figure 3-4 Candidate models for the neurodegenerative disease example

From discussions with the clinical team, the four candidate models displayed in Figure 3-4 were identified for the linear progression slopes; see Table 6-1 in the Appendix for the formal model definitions used in the following:

- Emax model with 90% of the maximum effect at 10 mg, corresponding to an ED_{50} = 1.11
- Quadratic model with maximum effect at 23 mg, corresponding to a standardized model parameter $\delta = -0.022$
- Exponential model with 30% of the maximum effect occurring at 20 mg, corresponding to a standardized model parameter $\delta = 8.867$
- Linear model

We use the simulated dataset to illustrate the MCP-Mod procedure in this situation, with an Emax dose response profile imposed on the linear slopes $\mu(d)$. The dataset is available via `data(neurodeg)` in the DoseFinding package. Figure 3-5 shows the simulated data per dose and patient.

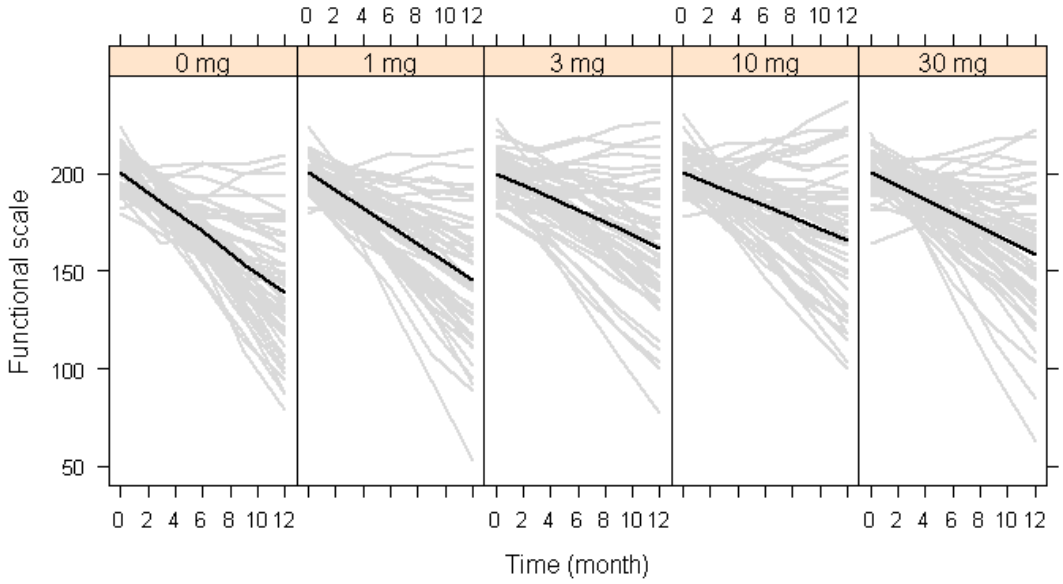

Figure 3-5 **Simulated data for the neurodegenerative disease example. Gray lines correspond to individual patient profiles, black line to a loess smoother.**

The **μ** vector of slopes is estimated via an LME fit of data, which can be done, for example, using the `lme` function in the `nlme` package in R, as illustrated below.

```
> data(neurodeg)
> head(neurodeg, n = 3)
      resp id dose time
1 191.7016  1    0    0
2 178.3995  1    0    3
3 167.3385  1    0    6
> fm   <- lme(resp ~ as.factor(dose):time, neurodeg, ~time|id)
> muH  <- fixef(fm)[-1]     # estimated slope
> covH <- vcov(fm)[-1,-1]   # var-cov of slopes
```

The estimated slopes for the simulated data are (–5.099, –4.581, –3.220, –2.879, –3.520)´. The corresponding estimated variance-covariance matrix **S** with compound symmetry structure has diagonal elements 0.149 and off-diagonal elements 0.009.

The optimal contrasts corresponding to the candidate models are calculated using the formula given in Section 3.1.2.2, with **S** given by the estimated variance-covariance matrix of $\widehat{\mu}$. The `DoseFinding` package includes the function `optContr` to calculate the optimal contrasts as

```
> # define candidate models
> doses <- c(0, 1, 3, 10, 30)
> mod <- Mods(emax = 1.11, quadratic=-0.022, exponential = 8.867, linear = NULL,
>             doses = doses)
```

```
> contMat <- optContr(mod, S=covH) # calculate optimal contrasts
```

The `MCTtest` function in the `DoseFinding` package implements the optimal model contrast tests for $\widehat{\mu}$ based on the multiple comparison approach described in Section 3.1.2.2. In the call below, `doses`, `mod`, and `optCg` are R objects representing respectively the doses, candidate models, and optimal contrasts for the study.

```
> MCTtest(doses, muH, S=covH, type = "general", critV = T, contMat=contMat)
. . .
Multiple Contrast Test:
            t-Stat  adj-p
emax         4.561 <0.001
quadratic    3.680 <0.001
linear       2.274  0.025
exponential  1.277  0.181


Critical value: 2.272 (alpha = 0.025, one-sided)
```

The quadratic and Emax model contrasts are significant, the linear model is borderline non-significant, and the exponential model clearly failed to reach significance at the 2.5% level. Therefore, the significance of a dose response signal is established and we can move forward to estimate the dose response profile and the target dose reaching the clinically relevant effect of 1.4.

Different approaches can be used for the dose response model fitting step: Generalized least squares fitting the estimates `muH` and `covH` or mixed-effects modeling (linear and nonlinear) incorporating a parametric dose response model for the progression slope μ(d). Both methods give very similar results; see Pinheiro et al. (2013). Here, we focus on generalized least squares fitting, which is implemented in the `fitMod` function in `DoseFinding`. We illustrate it in the call below for the Emax model, which is also the one with minimum AIC value.

```
> fitMod(doses, muH, S=covH, model="emax", type = "general")
Dose Response Model


Model: emax
Fit-type: general


Coefficients dose response model
     e0    eMax    ed50
-5.1808  2.1802  1.1873
```

A graphical description of the model fit, with the ANOVA estimates overlaid, can be found in Figure 3-6.

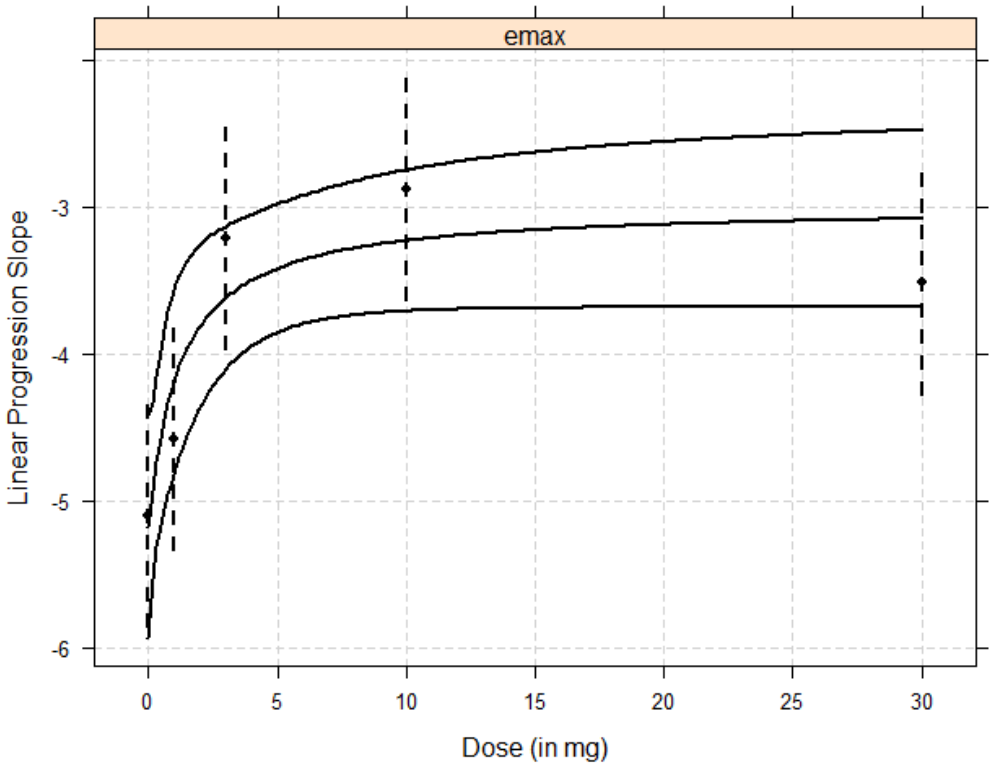

Figure 3-6 **Fitted Emax model with pointwise 95% confidence intervals and ANOVA estimates (black dots) from the linear mixed effects models with pointwise 95% confidence intervals.**

Estimates for the target dose, that is, the smallest dose producing an effect greater than or equal to the target value of 1.4, can be obtained via the `TD` function

In this example model selection was used to determine one dose response model to determine the dose response curve and the target dose. As mentioned earlier, model-averaging is an alternative approach if multiple models fit the data similarly well, we do not discuss it further here.

3.2 Results

In this section we provide evidence for the usefulness of the MCP-Mod approach described in Section 3.1. More specifically, we provide in Section 3.2.1 a list of medical papers that refer to MCP-Mod. In Section 3.2.2 we summarize the experiences at Novartis with the MCP-Mod procedure. We provide a list of Novartis studies that employed MCP-Mod and describe one study in detail, illustrating the use of MCP-Mod in practice and how it impacted the further development program. Finally, we summarize in Section 3.2.2 several numerical simulation studies that have been performed by the authors of this request as well as part of cross-industry initiatives to improve dose finding in clinical drug development.

3.2.1 List of medical papers

In the following, we provide a compiled list of references from the applied literature that used the MCP-Mod approach in their studies, as further evidence of its broad application.

- Calhoun, D. A. et al. (2011) Effects of a Novel Aldosterone Synthase Inhibitor for Treatment of Primary HypertensionClinical Perspective Results of a Randomized, Double-Blind, Placebo-and Active-Controlled Phase 2 Trial. Circulation 124.18: 1945-1955.
- Christiansen, S. et al. (2012). Mixtures of endocrine disrupting contaminants modelled on human high end exposures: an exploratory study in rats. International Journal of Andrology, 35, 303–316.
- Hass U. et al. (2007). Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. Environmental Health Perspectives, 115(suppl 1):122–128.
- Rosenstock K. J. et al. (2010). The 11-hydroxysteroid dehydrogenase type 1 inhibitor INCB13739 improves hyperglycemia in patients with Type-2 diabetes inadequately controlled by metformin monotherapy. Diabetes Care33:1516–1522, 2010.
- Ruilope L. M. et al. (2010). Blood-pressure reduction with LCZ696, a novel dual-acting inhibitor of the angiotensin II receptor and neprilysin: a randomised, double-blind, placebo-controlled, active comparator study. Lancet, 375, 1255-1266.
- Selmaj, K. et al. (2011, October). BAF312, a Selective Sphingosine 1-Phosphate Receptor Modulator, Effectively Suppresses MRI Lesion Activity in Relapsing-Remitting Multiple Sclerosis: Findings of an Adaptive Dose-Ranging Phase 2 Study. In Poster presented at the 5th Joint Triennial Congress of the European and Americas Committees for Treatment and Research in Multiple Sclerosis (Vol. 19, p. 22).
- Scholze M. and Kortenkamp A. (2007) Statistical Power Considerations Show the Endocrine Disruptor Low-Dose Issue in a New Light. Environ Health Perspect. 2007 December; 115(S-1): 84–90.
- So A. et al. (2010). Canakinumab for the Treatment of Acute Flares in Difficult-to-Treat Gouty Arthritis: Results of a Multicenter, Phase II, Dose-Ranging Study. Arthritis & Rheumatism, 62, 3064–3076.
- Verkindre C. et al. (2010) Sustained 24-h efficacy of NVA237, a once-daily long-acting muscarinic antagonist, in COPD patients. Respiratory Medicine Volume 104, Issue 10, October 2010, 1482–1489
- Villa, G. et al. (2011). Efficacy, safety, and tolerability of aliskiren monotherapy administered with a light meal in elderly hypertensive patients: a randomized, double-blind, placebo-controlled, dose-response evaluation study. The Journal of Clinical Pharmacology 52, 1901-1911

3.2.2 Experiences at Novartis

In this section we give an overview of several Novartis studies that employed MCP-Mod. We also describe one study in detail to illustrate the use of MCP-Mod in practice and how it impacted the further development program.

3.2.2.1 List of Novartis studies

In this section we provide a non-exhaustive overview of studies at Novartis, in which the MCP-Mod (or a conceptually closely related) statistical methodology has been used as a pre-specified methodology (in most cases as the primary analysis method, in a few cases as secondary or exploratory analysis for specific reasons). The list includes studies that are either completed or ongoing. Studies that are currently at the planning stage or where first patient first visit has not been achieved yet are not included.

Study ID	Phase	Condition studied	Treatment groups
ACZ885H2255	Phase IIb	Gout	5 doses, AC
ACZ885I2202	Phase IIb	Diabetes	PBO, 4 doses
ACZ885M2301	Phase III	Prevention of cardiovascular events	PBO, 3 doses
ACZ885M2301S1	Phase III	Prevention of cardiovascular events	PBO, 3 doses
ACZ885M2301S2	Phase III	Prevention of cardiovascular events	PBO, 3 doses
AEB071C2201	Phase IIb	Psoriasis	PBO, 3 od and 4 bid doses
BAF312A2201	Phase IIb	Multiple Sclerosis	PBO, 5 doses
BGG492A2207	Phase IIa/b	Epilepsy	PBO, 2 doses
LCI699A2201	Phase II	Hypertension	PBO, 3 od doses, 1 bid dose
LCQ908A2203	Phase IIb	Diabetes	PBO, 5 doses, AC
LCQ908B2302	Phase III	Familial Chylomicronemia Syndrome	PBO, 2 doses
LCQ908C2201	Phase II	Hypertriglyceridemia	PBO, 3 doses, 2 AC
LCZ696A2201	Phase IIb	Hypertension	PBO, 3 doses, 3 AC
LIK066A2202	Phase IIb	Diabetes	PBO, 7 doses
NVA237A2205	Phase IIb	COPD	PBO, 4 od doses, AC
NVA237A2208	Phase IIb	COPD	PBO, 3 bid doses, 4 od doses
QAW039A2206	Phase IIb	Asthma	PBO, 9 od doses, 4 bid doses, AC
QMF149B2201	Phase II	COPD	PBO, 4 doses
SAF312A2103	Phase IIa	Dental pain	PBO, 6 doses, AC
XBD179A2204	Phase II	Generalized anxiety disorder	PBO, 4 doses

Table 3-1 List of 20 example studies at Novartis which used MCP-Mod. PBO = Placebo; AC = Active Control

In Table 3-1 we list 20 examples studies in tabular form. In the individual paragraphs below we provide further details on the clinical background and how MCP-Mod was applied. Note that due to the broad applicability of the MCP-Mod approach, this list illustrates how MCP-

Mod can be tailored to the specific trial needs at hand. For example, in some cases the number of dose levels does not allow for dose response modeling (e.g. with 2 doses), in which case only the MCP part was used to test for a dose response signal. In other cases, testing for a dose response signal was not of major interest and interest focused on the Mod part only. The overarching principle of all of these examples is that multiple dose response models / shapes were pre-specified and either a test acknowledging for model uncertainty was used or model selection / model averaging was employed for dose response estimation. It also transpires from Table 3-1 that there is no limitation of using MCP-Mod in a specific indication or therapeutic area.

ACZ885H2255

Title: An adaptive dose-ranging, multi-center, single-blind, double-dummy, active-controlled trial to determine the target dose of canakinumab (ACZ885) in the treatment of acute flares in gout patients who are refractory or contraindicated to NSAIDs and/or colchicine.

Use of MCP-Mod: MCP-Mod used as primary analysis method to estimate the dose response relationship and to estimate the dose that has an efficacy comparable to active comparator.

ACZ885I2202

Title: Dose Finding, Safety and Efficacy of Monthly Subcutaneous Canakinumab Administration for the Treatment of Hyperglycemia in Metformin Monotherapy Treated Type 2 Diabetic Patients: a Randomized, Double-Blind, Placebo-Controlled, Multi-Center Study

Use of MCP-Mod: Planned to be used at interim to assess a dose response trend and choose a dose for final stage. The trial stopped at interim.

ACZ885M2301

Title: A randomized, double-blind, placebo-controlled, event-driven trial of quarterly subcutaneous canakinumab in the prevention of recurrent cardiovascular events among stable postmyocardial infarction patients with elevated hsCRP

Use of MCP-Mod: MCP-Mod used for exploratory assessment of the dose response curve for efficacy and safety outcomes.

ACZ885M2301S1

Title: A randomized, double-blind, placebo-controlled, event-driven trial of quarterly subcutaneous canakinumab in the prevention of recurrent cardiovascular events among stable post-myocardial infarction patients with elevated hsCRP: An imaging sub-study evaluating the effect of canakinumab on carotid atherosclerosis.

Use of MCP-Mod: This is a sub-study of ACZ885M2301. MCP-Mod is used to test for a dose response trend and to estimate the dose response relationship.

ACZ885M2301S2

Title: A randomized, double-blind, placebo-controlled, event-driven trial of quarterly subcutaneous canakinumab in the prevention of recurrent cardiovascular events among stable post-myocardial infarction patients with elevated hsCRP: An OGTT sub-study evaluating the effect of canakinumab on insulin secretion rate and other glucose control parameters following an oral glucose tolerance test in Type 2 diabetics

Use of MCP-Mod: This is a sub-study of ACZ885M2301. MCP-Mod is used to test for a dose response trend and to estimate the dose response relationship.

AEB071C2201

Title: A double blind, randomized, placebo controlled, multicenter, dose finding study of oral AEB071 assessing Psoriasis Area and Severity Index (PASI) response as a function of dose and treatment duration (primary outcome) in patients with plaque psoriasis.

Use of MCP-Mod: Primary analysis method to test for a dose response trend (using likelihood ratio tests) and estimate the dose response relationship and the target dose. Modeling of once and twice daily data has been performed by using a regimen multiplier.

BAF312A2201

Title: A phase II, double-blind, randomized, multi-center, adaptive dose-ranging, placebo-controlled, parallel-group study evaluating safety, tolerability, and efficacy on MRI lesion parameters and determining the dose response curve of BAF312 given orally once daily in patients with relapsing-remitting multiple sclerosis

Use of MCP-Mod: Two-stage dose-finding study. MCP-Mod used as primary analysis method to test for a dose response trend and estimate the dose response relationship. At interim the contrast test statistics were used to select the studied doses for the second part of the trial.

BGG492A2207

Title: A 12-week, multi-center, randomized, double-blind, placebo-controlled efficacy and safety study examining seizure frequency of BGG492 capsules administered orally three times daily (TID) as adjunctive treatment in patients with partial onset seizures.

Use of MCP-Mod: Primary analysis method to test for a dose response trend. No dose response modeling performed in this study as only two active doses were used.

LCI699A2201

Title: A multi-center, randomized, double-blind, placebo and active controlled, parallel group, dose finding study to evaluate the efficacy and safety of LCI699 compared to placebo after 8 weeks treatment in patients with essential hypertension

Use of MCP-Mod: Primary analysis for dose response analysis for od doses. For active comparator and the bid dose ANCOVA was used.

LCQ908A2203

Title: A 12-week multi-center, randomized, double-blind, placebo-controlled, parallel-group adaptive design study to evaluate the efficacy on blood glucose control and safety of five doses of LCQ908 (2, 5, 10, 15 and 20 mg) or sitagliptin 100 mg on a background therapy of metformin in obese patients with type 2 diabetes

Use of MCP-Mod:.Primary analysis method for testing for a dose response trend and estimating the dose response curve.

LCQ908B2302

Title: A randomized, double-blind, placebo controlled study to assess efficacy, safety and tolerability of LCQ908 in subjects with Familial Chylomicronemia Syndrome.

Use of MCP-Mod: MCP part of MCP-Mod used as primary analysis method to detect a dose response trend. This was a study in an orphan indication.

LCQ908C2201

Title: A multicenter, randomized, active comparator, placebo-controlled, double-blind pilot study to assess the efficacy and safety of LCQ908 alone and in combination with fenfibrate or Lovaza® in patients with severe hypertriglyceridemia

Use of MCP-Mod: Primary analysis method for dose response test and estimation of the dose response profile.

LCZ696A2201

Title: A multi-center, randomized, double-blind, placebo and active controlled, parallel group, dose range study to evaluate the efficacy and safety of LCZ696 comparatively to valsartan, and to evaluate AHU377 to placebo after 8 week treatment in patients with essential hypertension

Use of MCP-Mod: Planned as a secondary efficacy analysis

LIK066A2202

Title: A multi-center, randomized, double-blind, double-dummy, parallel group dose-finding study to evaluate the change in HbA1c after 12 weeks monotherapy with seven doses of LIK066 compared with placebo in patients with type 2 diabetes

Use of MCP-Mod: Primary analysis method to test for a dose response trend, estimate the dose response relationship and target doses of interest.

NVA237A2205

Title: A randomized, double-blind, placebo-controlled, 4 period incomplete block cross-over, multi-center, multiple dose (7 days) dose-ranging study to assess the efficacy and safety of 4 doses of NVA237 in patients with stable COPD, compared to seven days treatment with tiotropium (18µg once daily, open label) as an active control.

Use of MCP-Mod: Pre-specified as a supportive analysis.

NVA237A2208

Title: A randomized, double-blind, placebo-controlled, 2-period, cross-over study to assess the efficacy and safety of differing doses of NVA237 administered either once daily or twice daily, in patients with moderate to severe chronic obstructive pulmonary disease (COPD).

Use of MCP-Mod: Used as primary analysis method, but only modeling part was used in this study. Longitudinal dose response modeling has been used together with model averaging.

QAW039A2206

Title: A randomized, placebo-controlled, dose-ranging, multi-centre trial of QAW039 (1-450mg p.o.), to investigate the effect on FEV1 and ACQ in patients with moderate-to-severe, persistent, allergic asthma, inadequately controlled with ICS therapy

Use of MCP-Mod: Primary analysis method used to assess a dose response effect and estimate the dose response relationship. Modeling of once and twice daily data has been performed by using a regimen multiplier.

QMF149B2201

Title: A randomized, multi-center, parallel group, double blind, placebo and formoterol controlled 14 day dose ranging trial of 4 doses of indacaterol delivered via Twisthaler® in patients with COPD

Use of MCP-Mod: Primary analysis method used to assess a dose response effect, estimate the dose response relationship and estimate the target dose.

SAF312A2103

Title: A double-blind, randomized, single dose, placebo-controlled, three part study to evaluate the safety and tolerability of SAF312 in postoperative dental pain patients (Part A), to evaluate the analgesic effect of SAF312 in comparison to placebo in the treatment of postoperative dental pain using ibuprofen as a positive control (Part B) and to evaluate a dose response (Part C)

Use of MCP-Mod: Primary analysis method to estimate the dose response relationship. The design was adapted at the interim analysis between Part B and Part C based on the currently fitted dose response curve and the corresponding optimal design.

XBD173A2204

Title: A randomized, double-blind, placebo-controlled, parallel-group study of the efficacy, safety and tolerability of XBD173 in patients with generalized anxiety disorder (GAD).

Use of MCP-Mod: Primary analysis method to assess a dose response trend and estimate the dose response relationship.

3.2.2.2 Case Study: NVA237A2205 study

Chronic Obstructive Pulmonary Disease (COPD) is a disease of the lungs characterized by airflow limitation which is not fully reversible. The investigational drug NVA237 is a dry

powder formulation of the muscarinic receptor antagonist glycopyrronium bromide being developed by Novartis.

The primary purpose of the A2205 study was to provide data about the risk-benefit of four doses of NVA237 (12.5, 25, 50 and 100µg o.d.) and open-label tiotropium (18µg) so that an optimal dose of NVA237 can be chosen for Phase III studies. The primary endpoint of the study was to evaluate the bronchodilatory efficacy of NVA237 in patients with stable COPD in terms of trough forced expiratory volume over 1 second (FEV1) (mean of 23h 15min and 23h 45min post dose) in Liter (L) following 7 days of treatment.

Facilitated by the relatively quick read-out and to reduce the impact of inter-patient variability, which is known to be large for this endpoint, a crossover design was used. The relatively large number of treatments (6) and the required wash-out period (7 days) between treatments made a complete crossover unfeasible and a balanced incomplete block crossover design was used, in which each patient received 4 of the 6 treatments. The treatment sequences, in blocks of size 30, were determined prior to randomizing patients, so as to ensure balance in the order and combinations of treatments measured in the same patient. In this study, the pre-planned primary analysis was an analysis-of-covariance (ANCOVA). The sample size was thus determined to achieve a specified power for the ANCOVA pairwise tests, using a conservative Bonferroni correction to adjust for multiplicity. MCP-Mod was specified for the supportive analysis, but ultimately played a key role in the final discussions on which dose to take forward into Phase III.

Patients were accounted for as fixed effects in the dose response model, together with treatment sequence, period and baseline FEV1. We now consider each of the five MCP-Mod steps described in Figure 3-1 for this particular trial; see also Bretz et al. (2009) for more details on this analysis.

Step 1: Set of candidate models

After discussions with the clinical team, five candidate models were selected to represent the anticipated dose response shapes for the improvement in FEV1 change from baseline over placebo: Emax (2 shapes with ED_{50} = 2.65 and ED_{50} = 12.5, respectively), linear, logistic (with ED_{50} = 29 and δ = 9.55), and quadratic (with δ = -0.0075). The candidate model shapes were elicited through discussions with the clinical team and utilizing results from previous studies on the same compound and other drugs for the same indication. It was anticipated that the maximum treatment effect would be a 0.15 L improvement over placebo. The resulting candidate models are illustrated in Figure 3-5.

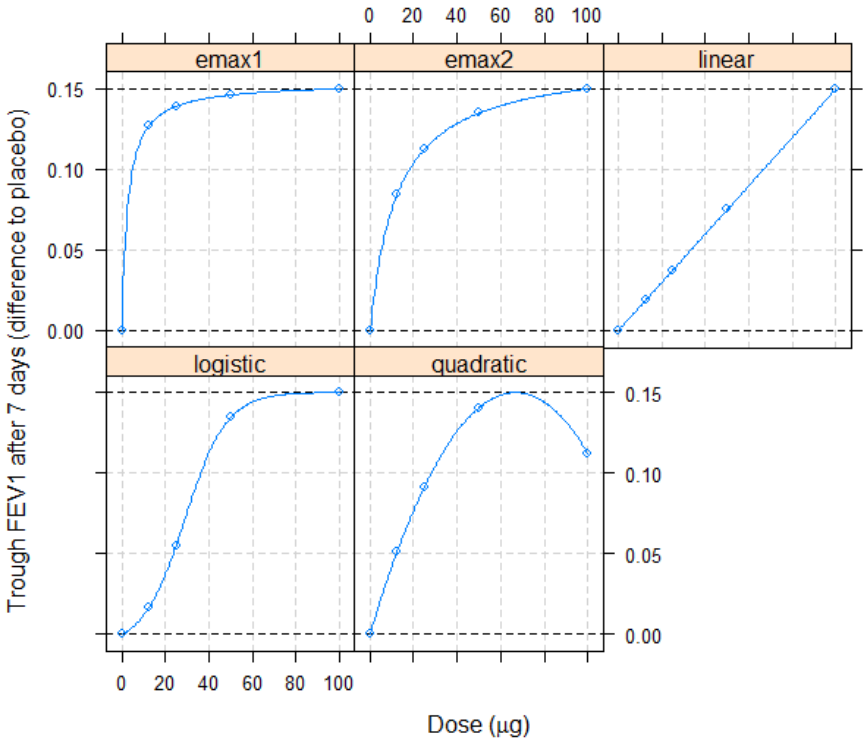

Figure 3-5 Candidate models specified before trial start.

Step 2: Optimal model contrasts

Because of the presence of covariates, the covariance matrix of the treatment effect estimates depends on covariates through the design matrix and hence the general formula given in Section 3.1.2.2 needs to be used to determine the optimal contrasts. The design matrix used in this step was obtained prior to the start of trial, depending only on the subset of covariates known at that stage, namely: patient, period, and dose.

Step 3: Testing for dose response signal

From this step onward, the real data observed in the trial is used to illustrate the methods and derive results. Applying the optimal contrasts to the treatment estimates, one obtains that all contrasts had test statistics > 6 and multiplicity adjusted p-values < 0.0001. As a result, the significance of the dose response signal was established and all models were considered in the next step.

Step 4: Model selection

The AIC criterion was used to select the best model. Note that, even though there are two Emax shapes in the candidate set, only one Emax fit is obtained. Based on the AIC results, the Emax model was chosen to represent the dose response profile. The estimated improvement over placebo estimated from the fitted Emax model was $1.69d/(18 + d)$.

Step 5: Dose estimation

Based on the fitted Emax of Step 4, the smallest dose giving the clinically relevant improvement over placebo of 0.12 L is estimated to be 44 µg. This is the MED estimate produced by MCP-Mod in this study. The precision of the MED estimate was evaluated via a bootstrap approach: The 90% confidence interval for the MED, corresponding to the 5% and 95% quantiles of the bootstrap sample, was [18, 81], reflecting the uncertainty in the estimate. Figure 3-6 displays the fitted model and corresponding confidence intervals.

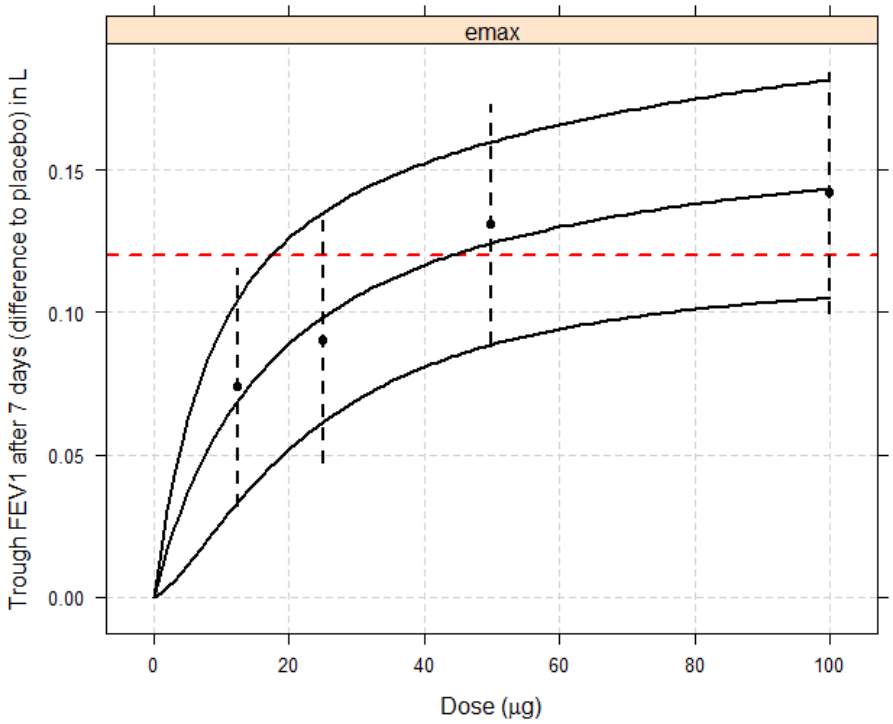

Figure 3-6 **Fitted Emax model with pointwise 95% confidence intervals for the NVA237A2205 study data and ANCOVA estimates (black dots). The red dashed line is the clinical relevance threshold of 0.12 L.**

Although MCP-Mod was not pre-specified as primary analysis method in this study, it provided answers to both key questions that the clinical team wanted to address: establishing a dose response signal and selecting a dose for the confirmatory phase. The former was unequivocally shown by the large test statistics for the model contrasts, while the latter was provided by modeling the dose response profile. Relying on the original primary analysis methods based on ANCOVA pairwise comparisons, the quantitative basis for the decisive discussions would have been considerably weaker.

The results of this analysis combined with additional clinical considerations finally suggested the use of the 50 µg once daily dose in the main pivotal studies A2303 and A2304, which led to a positive outcome. The drug was granted market authorization with the 50 µg o.d. dose in Europe in 2012; see EMA document EMA/CHMP/508029/2012.

3.2.3 Simulation studies

To allow a direct quantitative assessment of the MCP-Mod approach, we describe in this Section two comprehensive simulation studies, each motivated by a real dose-ranging applications and covering a wide range of practical scenarios. First, we summarize the simulation results from Branson et al. (2003), who primarily compared the MCP-Mod approach with competing trend tests. Second, we review the simulation work performed by the PhRMA ADRS working group, which investigated several innovative approaches (including MCP-Mod and a response-adaptive version) in two extensive simulation studies aimed at quantifying the benefits of ADR methods over traditional, fixed design approaches (Bornkamp et al. 2007; Dragalin et al. 2010). For each simulation study, we describe its design, including its assumptions and scenarios; the performance metrics used to evaluate different statistical operational characteristics of each method; and a graphical summary of the statistical performance of the methods, based on the simulation results.

3.2.3.1 Simulation study by Branson et al. (2003)

In this section we investigate, via simulation, the performance of the MCP-Mod dose finding approach with respect to two main aspects:

- its power to detect the existence of a dose response signal, and
- its ability to, at the end, choose a dose close to the desired level (taking into account both statistical significance and clinical relevance), the dose selection performance.

Other classical dose finding methods based on multiple testing procedures were also used in the simulations for comparison with the MCP-Mod approach, with regard to the dose response signal detection performance. Because model-based dose selection methods can choose any value on a continuous scale, they are not directly comparable to classical dose finding methods based on multiple comparisons alone, which can only select the dose from within the set of levels under investigation. Therefore, the dose-selection performance is only investigated for the MCP-Mod approach.

Design

The study design used for the simulations was based on the following assessments, matching the requirements of the original MCP-Mod procedure described in Section 3.1:

- dose levels: d = 0, 0.05, 0.2, 0.6 and 1
- five parallel groups, with a single endpoint measured per patient
- assumptions on endpoint values at dose d: independently distributed as $N(\mu(d); \sigma^2)$
- balanced sample size allocation with n patients per dose group and no drop-outs
- group sample sizes: n = 10, 25, 50, 75, 100 and 150
- one-sided significance level $\alpha = 0.05$

Table 3-2 lists the dose response models included in the simulation study. All of these shapes have the property that at d = 0 the response value is about 0.2 and, with the exception of the constant shape, all have a maximum response of about 0.8 within the interval [0, 1] (that is, a

maximum dose effect of about 0.6). Figure 3-7 displays the dose response profiles for the nine shapes listed in Table 3-2. A total of 10,000 simulated trials were generated for each combination of shape and sample size.

The constant shape is included to evaluate the performance of the MCP-Mod method in terms of preserving the nominal Type I error rate for dose response signal detection. Shapes 2 through 7 form the set of candidate models for the contrast tests. The last two shapes, 8 and 9, are included to evaluate the performance of the MCP-Mod approach under model misspecification: They do not quite correspond to any of the model shapes in the candidate set, though can be approximated by some of the models in there.

Model	**μ(d)**
constant	0.2
Emax	$0.2 + 0.7d/(0.2 + d)$
linear in log-dose	$0.2 + 0.6\ \log(5d + 1)/\log(6)$
Linear	$0.2 + 0.6d$
exponential	$0.2\exp[\log(4)d]$
quadratic	$0.2 + 2.0485d - 1.7485d^2$
logistic	$0.193 + 0.607/\{1 + \exp[10\log(3)(0.4 - d)]\}$
double-logistic	$\{0.198 + 0.61/(1 + \exp[18(0.3 - d)])\}\ I(d \leq 0.5) + \{0.499 + 0.309/(1 + \exp[18(d - 0.7)])\}\ I(d > 0.5)$
convex	$0.2 + 0.6/\{1 + \exp[10(0.8 - d)]\}$

Table 3-2 Dose response models included in the simulation study

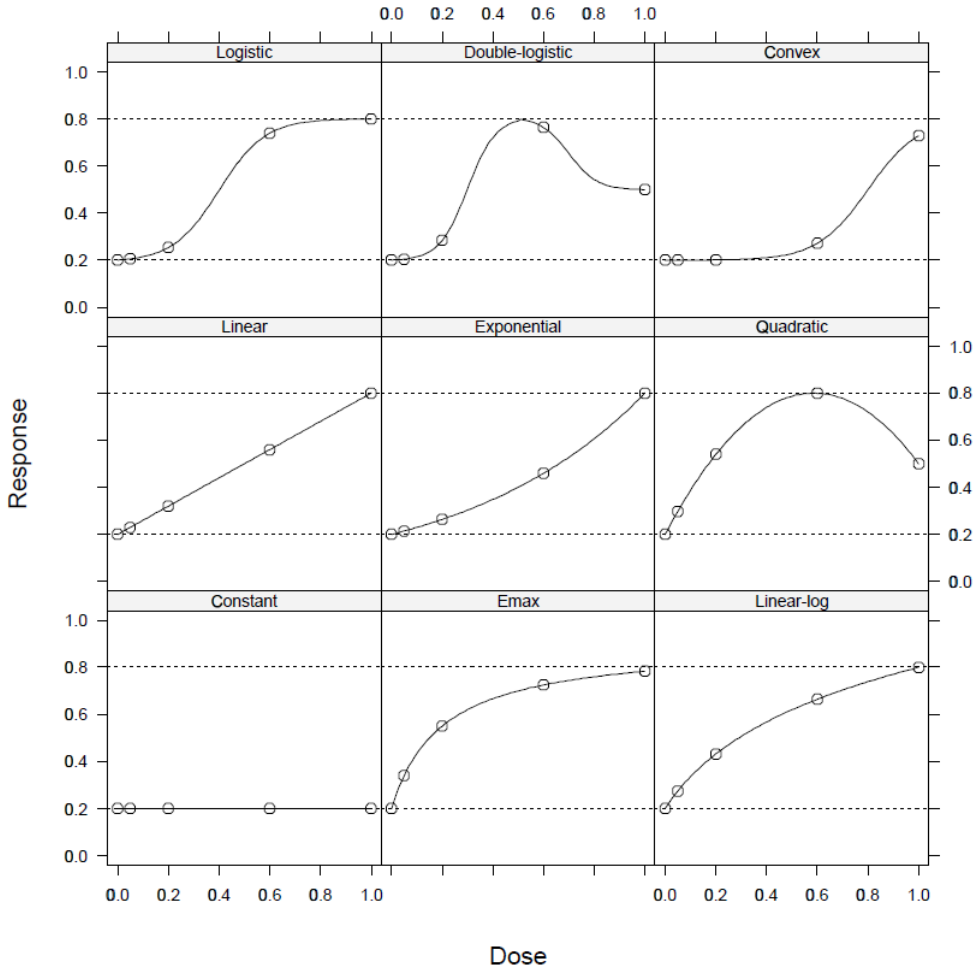

Figure 3-7 Visualization of the dose response models included in the simulation study. Open dots indicate the responses at the selected dose levels.

Dose response signal detection performance

For the purpose of evaluating the dose response signal detection performance of the MCP-Mod method, the response standard deviation was set at $\sigma = 1.478$, which, for a sample size of $n = 75$ patients per arm, gives a power of 80% for the pairwise test between two doses at the maximum effect of $\delta = 0.6$. Table 3-3 gives the simulated probabilities of establishing a dose response signal for the different methods under the various combinations of shapes and sample sizes. We included the likelihood ratio test (LRT; Bartholomew, 1961) and the step contrasts (Bauer and Hackl, 1985; Ruberg, 1989) as competitors to the MCP-Mod approach. The LRT is known to be one of the most powerful tests for trend throughout the order restricted alternative region $\mu_1 \leq \ldots \leq \mu_k$. In contrast to MCP-Mod, the LRT is designed to test only for a dose response trend and thus does not give any information about the underlying dose response shape. The step contrasts are a powerful alternative to the LRT. The step contrasts are particularly powerful for finding the change point in a series of treatment

means (Bauer and Hackl, 1985). For our simulations we used the multivariate t-distribution to compute the critical values.

		Data generating shape								
Method	**n**	**Const**	**Emax**	**Lin-log**	**Linear**	**Exp**	**Quad**	**Logistic**	**D.logist**	**Convex**
MCP-Mod	10	0.046	0.248	0.261	0.245	0.241	0.219	0.317	0.223	0.182
	25	0.049	0.47	0.491	0.484	0.461	0.389	0.599	0.411	0.337
	50	0.048	0.72	0.752	0.734	0.712	0.642	0.856	0.66	0.554
	75	0.051	0.868	0.891	0.88	0.862	0.799	0.96	0.805	0.728
	100	0.049	0.944	0.952	0.949	0.942	0.896	0.989	0.901	0.848
	150	0.052	0.989	0.992	0.992	0.988	0.972	0.999	0.976	0.952
LRT	10	0.048	0.242	0.253	0.246	0.245	0.18	0.283	0.181	0.196
	25	0.052	0.46	0.475	0.468	0.464	0.311	0.555	0.328	0.359
	50	0.051	0.713	0.731	0.718	0.706	0.523	0.823	0.556	0.591
	75	0.052	0.857	0.874	0.865	0.858	0.684	0.946	0.721	0.761
	100	0.048	0.937	0.943	0.941	0.935	0.806	0.982	0.838	0.869
	150	0.049	0.987	0.99	0.989	0.987	0.925	0.999	0.947	0.965
step	10	0.048	0.229	0.24	0.229	0.234	0.177	0.272	0.179	0.19
	25	0.051	0.44	0.452	0.447	0.444	0.312	0.539	0.329	0.354
	50	0.053	0.69	0.711	0.697	0.687	0.52	0.818	0.559	0.58
	75	0.055	0.841	0.858	0.85	0.844	0.682	0.944	0.726	0.753
	100	0.048	0.926	0.933	0.933	0.927	0.804	0.981	0.839	0.865
	150	0.049	0.984	0.988	0.987	0.983	0.923	0.999	0.95	0.962

Table 3-3 Dose response signal detection probabilities for different trend tests, under the shape and sample size combinations

The Type I error rate is well controlled at its nominal level of 5% for all sample sizes. The logistic shape has the highest power and the convex shape the smallest. The power values for the other shapes are of comparable magnitude. In particular, it is seen that MCP-Mod behaves very similar in power compared with the LRT. The LRT is slightly better for the convex shape. This is because we did not include the convex shape as a model in the candidate set for MCP-Mod. Had such model been included in the candidate set, the advantage of the LRT would likely vanish. Instead, MCP-Mod seems to be slightly more powerful for the linear-log, linear and logistic shapes. MCP-Mod is considerably more powerful than the LRT for the quadratic and the double-logistic shapes, since the LRT is not designed for such downturns at higher doses. Both the MCP-Mod and the LRT are more powerful than the step contrasts.

Dose selection performance

For the purpose of evaluating the dose selection performance of the MCP-Mod approach, we use a smaller, more realistic response standard deviation of at $\sigma = 0.65$ (Branson et al., 2003).

The clinically relevant effect was set at $\Delta = 0.4$, i.e. we focus on estimating the smallest dose that achieves a target difference of 0.4 on top of the placebo response.

We omit reporting the simulation power values under the different shape scenarios for the new σ. As expected, the power values are considerably larger than in Table 3-3, because of the 56% reduction in σ. The Type I error rate is well controlled at the nominal 5% level and by $n = 50$ all shapes result in almost certain dose response signal detection.

The dose response signal detection results also provide information about the ability of the contrast tests in the MCP-Mod approach to discriminate between the models in the candidate set. Table 3-4 gives the simulation probabilities of choosing the correct model for the six models in the candidate set (e.g., the probability of the Emax model contrast yielding the largest contrast test statistic when in fact this is the correct model). The quadratic model has the best discrimination power, since its associated contrast is the least correlated with the remaining model contrasts. The linear-log and linear model are the hardest to identify, which again can be explained through their high correlation. Because models which can represent similar dose-response profiles will likely lead to similar dose selections in the second stage of the method, the discrimination among highly correlated models is less critical than among the less correlated ones. This issue will be further explored below.

n	Emax	Lin-log	Linear	Exp	Quad	Logistic
10	0.28	0.09	0.08	0.38	0.45	0.41
25	0.5	0.22	0.18	0.59	0.79	0.63
50	0.66	0.34	0.31	0.67	0.91	0.75
75	0.73	0.44	0.38	0.71	0.96	0.83
100	0.77	0.54	0.47	0.75	0.98	0.87
150	0.84	0.65	0.59	0.79	0.99	0.93

Table 3-4 Probability of correctly identifying the response model, for the six models in the candidate set and the different sample sizes for σ = 0.65

Table 3-5 gives the target doses to achieve the desired clinically relevant effect of 0.4 (difference with respect to placebo) for the eight different shapes considered for dose selection. Due to the large number of shape and sample size combinations, we only present the dose selection simulation results for a subset of the scenarios investigated. We estimate the target doses using three different methods A, B, and C, using the upper confidence bound, the mean estimate or the lower confidence bound, respectively; see the definition of MED1, MED2, and MED3 in Figure 3-8.

Shape	Emax	Lin-log	Linear	Exp	Quad	Logistic	D.logist	Convex
Target dose	0.27	0.46	0.67	0.79	0.25	0.46	0.34	0.87

Table 3-5 Target doses for clinically relevant effect of 0.4 under various simulations shapes

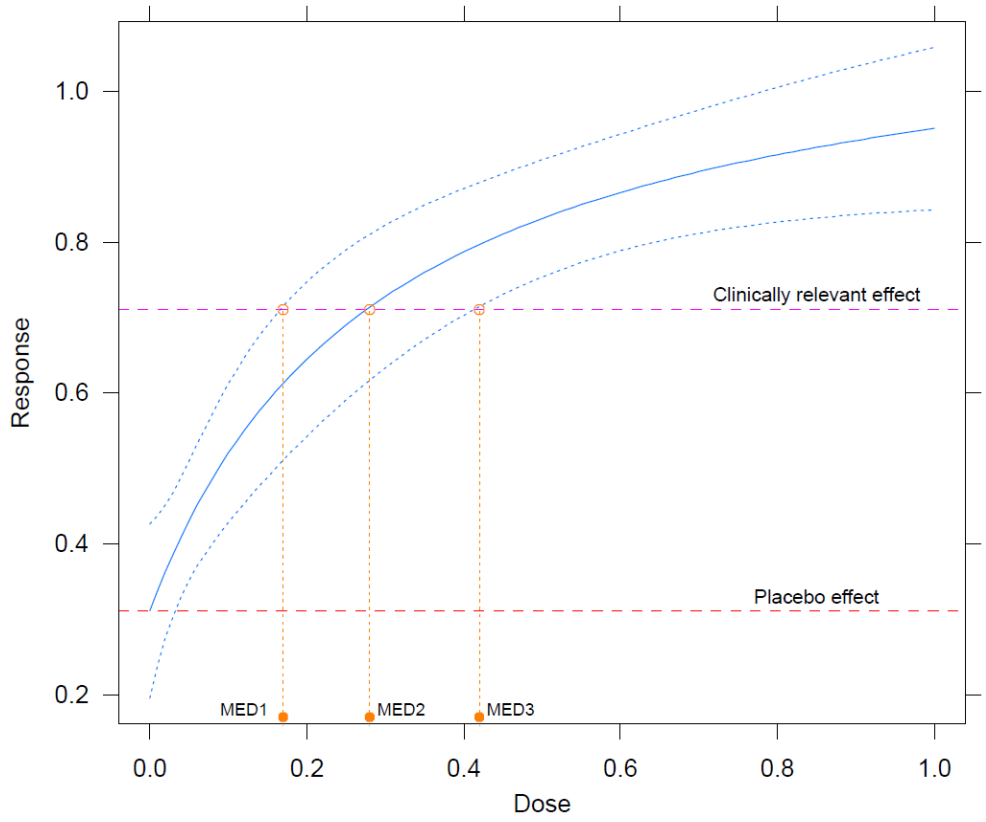

Figure 3-8 Three different methods for target dose estimation

Figures 3-9 and 3-10 include the boxplots of the selected doses in 10,000 simulated trials using three methods A, B, and C to estimate the target dose for all shapes and n = 25 and 75, using pointwise 80% confidence intervals. It is clear from the figures that method A tends to underestimate the target dose, method C tends to overestimate it, and method B estimates the target dose more consistently. The precision of the methods is considerably enhanced when the sample size increases from 25 to 75. It should be noted, though, that a more suitable choice of doses with modeling in mind would yield considerably better results, for the same overall sample sizes. This issue was considered extensively by the PhRMA ADRS working group, whose simulation results (including the MCP-Mod approach) are considered in the next subsection.

We conclude this section with the remark that the precision of the dose selection algorithms vary considerably with the underlying dose response shape. The quadratic, convex and exponential shapes tend to lead to greater precision (for the particular scenarios used here, but not in general), while the remaining shapes give similar dispersions for the dose estimates (except for the linear-log shape, which gave higher dispersion than the others).

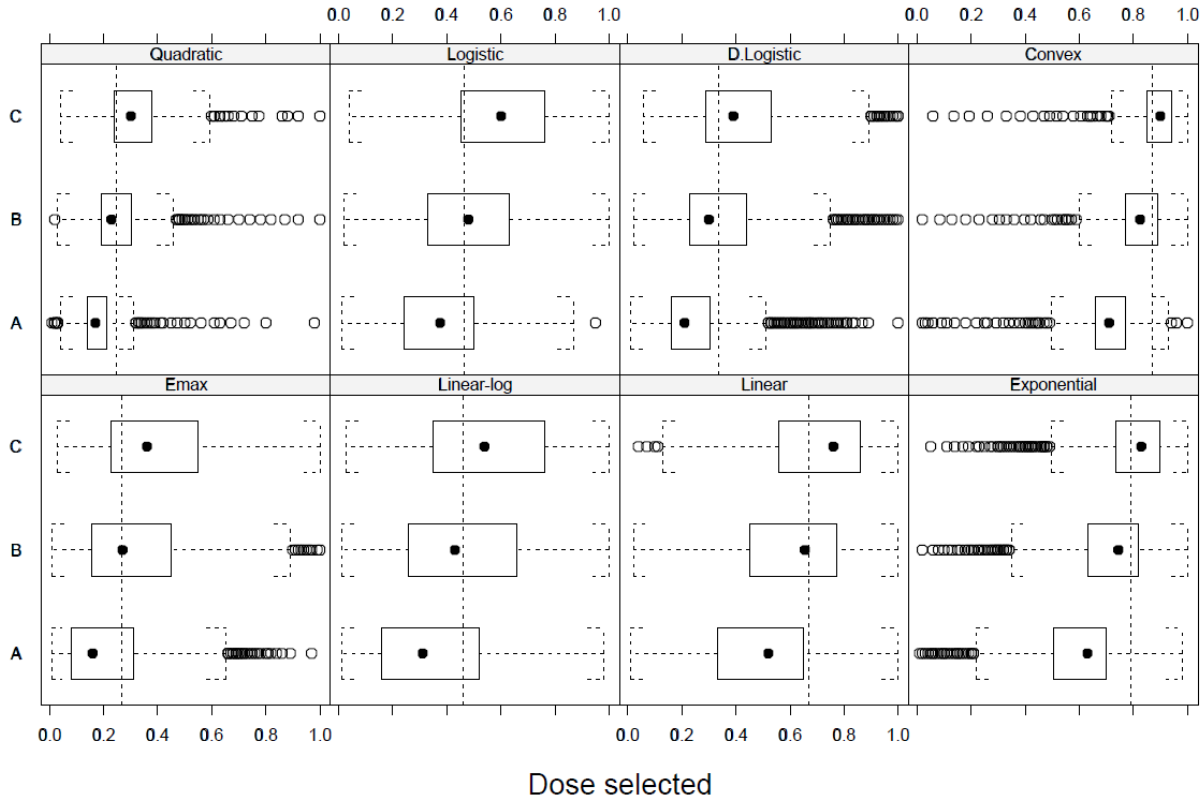

Figure 3-9 Boxplots of selected doses for n = 25

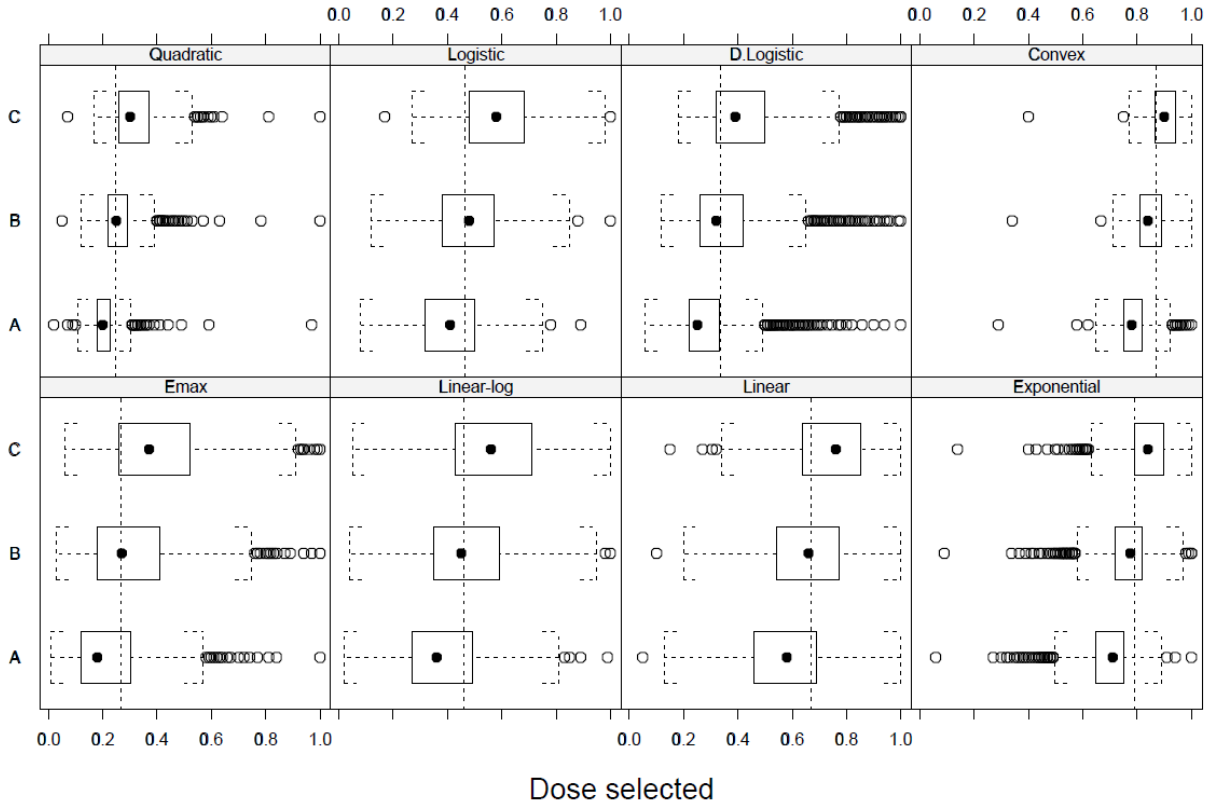

Figure 3-10 Boxplots of selected doses for n = 75

3.2.3.2 PhRMA ADRS working group simulations

The simulation results in this section are taken from the white paper of the PhRMA working group on "Adaptive Dose-Ranging Studies" (Bornkamp et al. 2007). The main objective of this group was to evaluate different novel and existing non-adaptive and adaptive dose-ranging methods in a comprehensive simulation study. A variant of MCP-Mod was one of the evaluated methodologies and a subset of the performed simulations is presented here. The advantage of these simulations is that they allow comparing MCP-Mod to alternative methodologies.

For conciseness and comparability, the comparison is limited here to methods that were fixed-design methods since MCP-Mod was employed in a fixed design setting in these simulations; see Dragalin et al. (2010) for a comparison of an adaptive version of MCP-Mod to other adaptive approaches. In addition, we removed one additional method that was conceptually very similar to MCP-Mod. The remaining comparators then include a Bayesian, a nonparametric and an ANOVA approach, the latter being a benchmark for the alternative methods.

Compared methodologies

Below we give descriptions of the competing methodologies and how they were applied in the simulation study. MCP-Mod was employed here with model-selection (instead of model-averaging) based on the maximum contrast test statistics.

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

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

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

Design of simulation study

In total, 3 x 6 x 2 = 36 different scenarios were used in the simulations, corresponding to different combinations of dose designs (3), dose-response profiles (6), and total sample size (2). To give practical motivation, a neuropathic pain dose-ranging study was used to provide context for the simulation study. The primary endpoint is the change from baseline in a visual

analog scale (VAS) at 6 weeks and the response is assumed to be normally distributed with variance 4.5. Negative values give indication of efficacy in reducing the neuropathic pain. The clinically relevant effect is set to $e_{targ} = -1.3$ units (i.e., an average reduction of at least 1.3 units from baseline).

It is assumed that up to nine equally spaced doses can be utilized in the trial: 0, 1, …, 8. Three different dose designs are considered in the simulations, to investigate the impact of number and spacing of doses on the performance of the methods:

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

Figure 3-11 displays a total of five different dose-response profiles that were used to simulate the primary endpoint, allowing the evaluation of the methods under a wide range of scenarios likely to be observed in clinical practice. A flat model was also included to evaluate the Type I error rate. In all models, the placebo effect was set to 0 points and, with the exception of the logistic model, the maximum effect within the observed dose range was set to -1.65 units.

- Flat: $\mu(d) = 0$
- Linear: $\mu(d) = -(1.65/8)d$
- Logistic: $\mu(d) = 0.015 - 1.73/\{1 + \exp(1.2/(4 - d))\}$
- Umbrella: $\mu(d) = -(1.65/3)d + (1.65/36)d^2$
- Emax: $\mu(d) = -1.81\ d/(0.79 + d)$

The response at dose d are assumed to be independently normally distributed with mean $\mu(d)$ and variance 4.5.

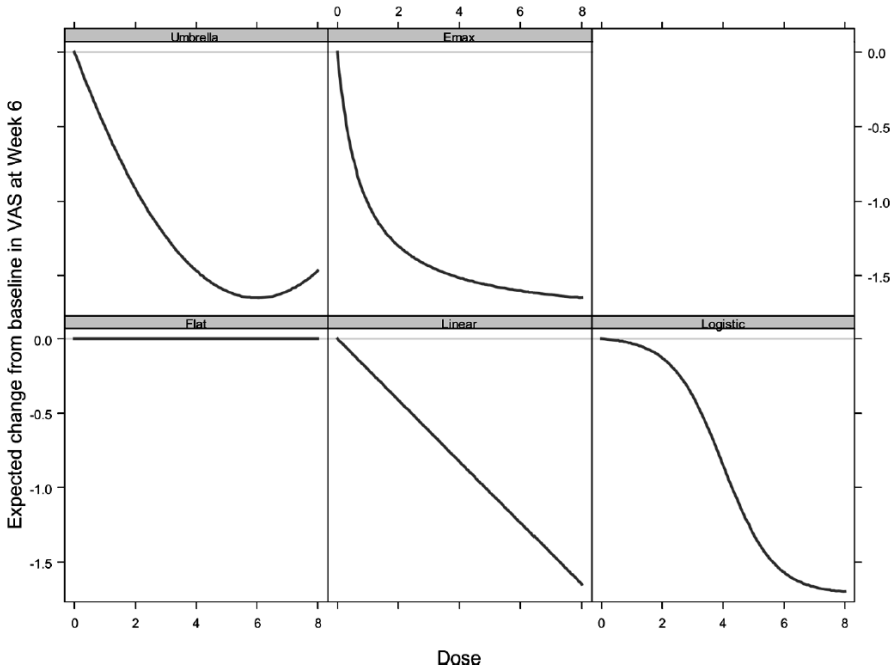

Figure 3-11 Dose-response profiles used in simulation

Two maximum total sample sizes were used in the simulations: 150 and 250 patients. These values are consistent with sample sizes commonly used in neuropathic pain Phase II trials. The total sample size N corresponds to the sum of the number of patients assigned to each dose. For example, under equal treatment allocation and with N = 250, a 5-dose design assigns 50 patients to each dose and a 9-dose design, about 28. To adequately estimate the statistical operational characteristics of the various methods, a minimum of 5,000 simulated trials were used for each of the scenarios considered. For some of the less computationally intensive methods (MCP-Mod and ANOVA), 10,000 simulated trials per scenario were used.

Measuring Performance of Methods

To make the problem more concrete, for the purpose of the simulation study the following specific goals were identified:

(A) Detecting dose response: evaluate if there is evidence of activity associated with the drug, represented by a change in clinical response resulting from a change in dose;

(B) Identifying clinical relevance: if a significant trend is established, determine if a predefined clinically relevant response can be obtained within the observed dose range;

(C) Selecting a target dose: when the previous goal is met, select the dose to be brought into the confirmatory phase, the so-called target dose;

(D) Estimating the dose response: finally, estimate the dose-response profile within the observed dose range.

Performance metrics to quantify each of these goals are described below.

(A) Detecting dose response. Each of the methods includes a decision rule to determine whether the data provides sufficient evidence of a dose response (DR) signal. The probability of identifying the presence of dose response, P(DR) estimated as the percentage of simulated trials in which the decision rule concluded for dose response signal is used as the summary metric for this objective. To allow adequate comparisons, a significance level of 5% was specified for all methods.

(B) Identifying clinical relevance. It is, of course, possible to conclude that dose response is present, but, nevertheless, that none of the observed doses is capable of producing at least the clinically relevant effect. All methods implement decision rules for identifying clinical relevance within the dose range of the trial. The corresponding probability, Pr(dose), estimated as the percentage of simulated trials in which a significant trend was established and a dose with a clinically relevant effect was chosen, is used to summarize the performance of the methods with regard to this objective. By definition, Pr(dose) ≤ Pr(DR).

(C) Selecting a target dose. In practice, the selection of the dose to bring into the confirmatory phase is based on a plurality of factors, including, but not restricted to, efficacy and safety outcomes in the Phase II trial(s). For the purpose of this simulation study the problem was simplified and only a target efficacy result (the clinically relevant effect) is used to determine the dose to be selected. In this context, the target dose d_{targ} is defined as the smallest dose which produces an effect greater than, or equal to, the clinically relevant target effect e_{targ}. For the purpose of the simulation study the dose selection is restricted to the set {1, 2, …, 8}. Therefore, estimated target doses resulting from any of the model-based methods are rounded to the nearest integer within this set. Note that this may result in a dose not used in the trial being selected in the end as the target dose (e.g., d = 3 being chosen in the 5-dose design). Table 3-5 lists the values of d_{targ}.

The distribution of estimated d_{targ}, from the simulated trials, provides a complete description of the performance of the estimate. The following statistics were used to summarize the dose estimation performance of the various methods, with expectations and probabilities referring to the corresponding Monte Carlo distributions obtained in the simulations.

- Percentage bias:

$$pBias = 100E(\hat{d}_{targ} - d_{targ})/d_{targ}$$

- Percentage absolute error:

$$pError = 100E(|\hat{d}_{targ} - d_{targ}|)/d_{targ}$$

(D) Estimating the dose response. To characterize the efficacy of the compound precisely proper estimation of the dose response profile is necessary. The average absolute prediction error (APE), calculated at the available doses (including placebo), is used as an overall measure of performance for dose response estimation. Letting μ(d) denote the expected dose response at dose d and $\widehat{\mu}(d)$ its prediction based on the estimated dose-response model, we define APE = $1/9\Sigma_d$ E($|\widehat{\mu}(d) - \mu(d)|$). To make the summary statistic non-dimensional and interpretable as a percentage, we consider the percent APE (pAPE), defined as the percent value of APE relative to the (absolute) target effect, that is pAPE = $100APE/e_{targ}$.

n	Actual	Rounded
Linear	6.30	6
Logistic	4.96	5
Umbrella	3.24	3
Emax	2.00	2

Table 3-6 Target doses and target dose intervals for dose response models used in the simulation

Simulation Results

Below we present a comprehensive subset of the simulations; see Bornkamp (2007) for more extensive results.

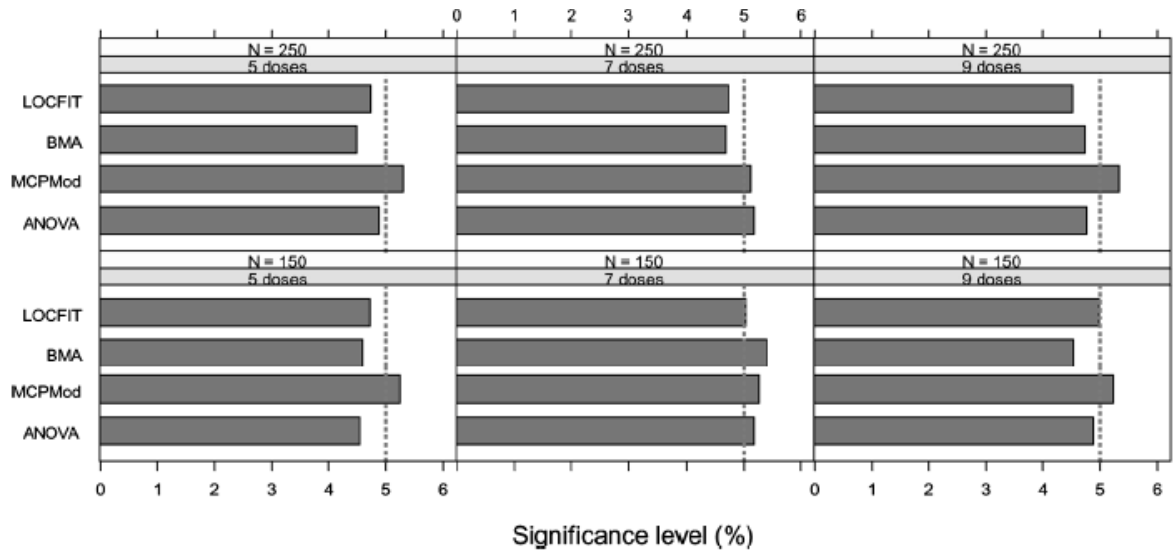

Figure 3-12 Type I error rate for detecting dose response under flat profile

(A) Detecting dose response. As illustrated in Figure 3-12, all methods were capable of controlling the Type I error rate. Fluctuations around the 5% level are consistent with Monte Carlo error. The probabilities of detecting dose response under active profiles are included in Figure 3-13. For a total sample size of N = 250 patients, MCP-Mod and the other methods have reasonable power to detect dose response under the different dose response models. ANOVA presents the relatively worst performance compared to the remaining methods in particular for the case of 9 doses, because it adjusts for multiplicity for the individual dose-comparisons. When the sample size is reduced to N = 150 patients, the differences among the methods become more pronounced.

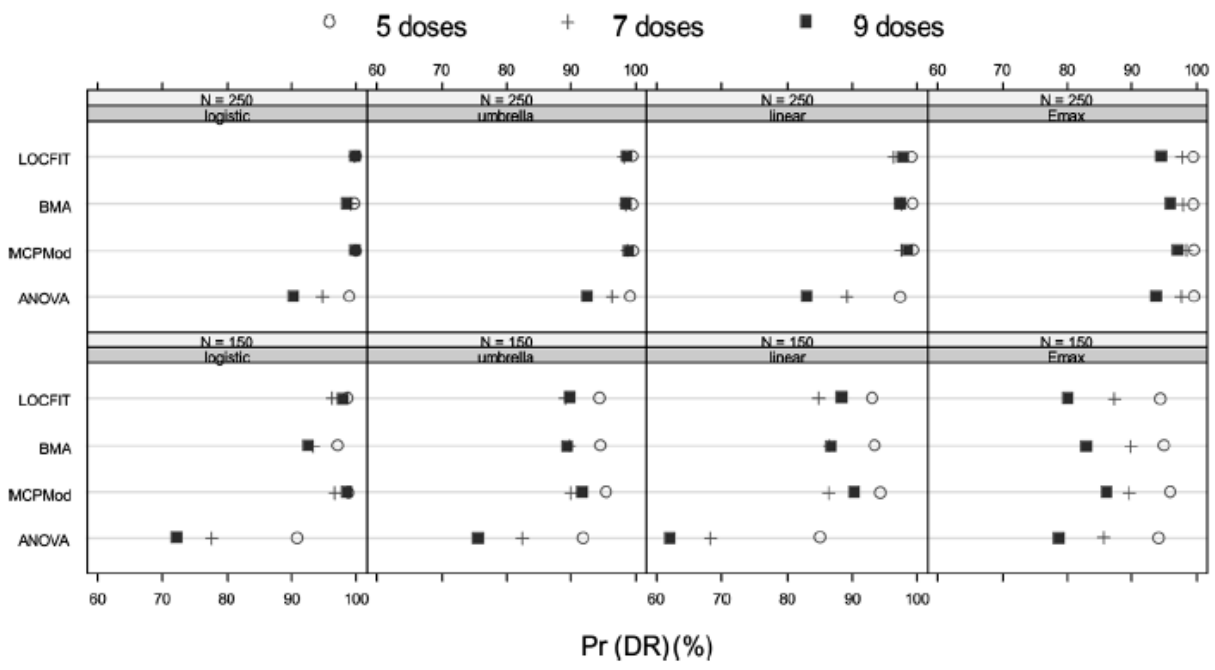

Figure 3-13 Power to detect dose response under active DR profiles.

(B) Identifying clinical relevance. Figure 3-14 shows the probabilities of incorrectly identifying a dose that produces a clinically relevant effect under a flat dose response model. ANOVA has the worst performance for this metric. This is due to the fact that no smoothing is done between the dose levels and random highs or random lows influence the dose-estimation for the ANOVA approach. The performance deteriorates further with the increase in number of doses and reduction in sample size. The probabilities of correctly choosing a clinically relevant dose under active dose response profiles are presented in Figure 3-15. The performance of the methods varies considerably with the underlying dose response model, the total sample size, and the dose design. None of the methods clearly dominates the others and MCP-Mod is comparing quite favorably to the other methods.

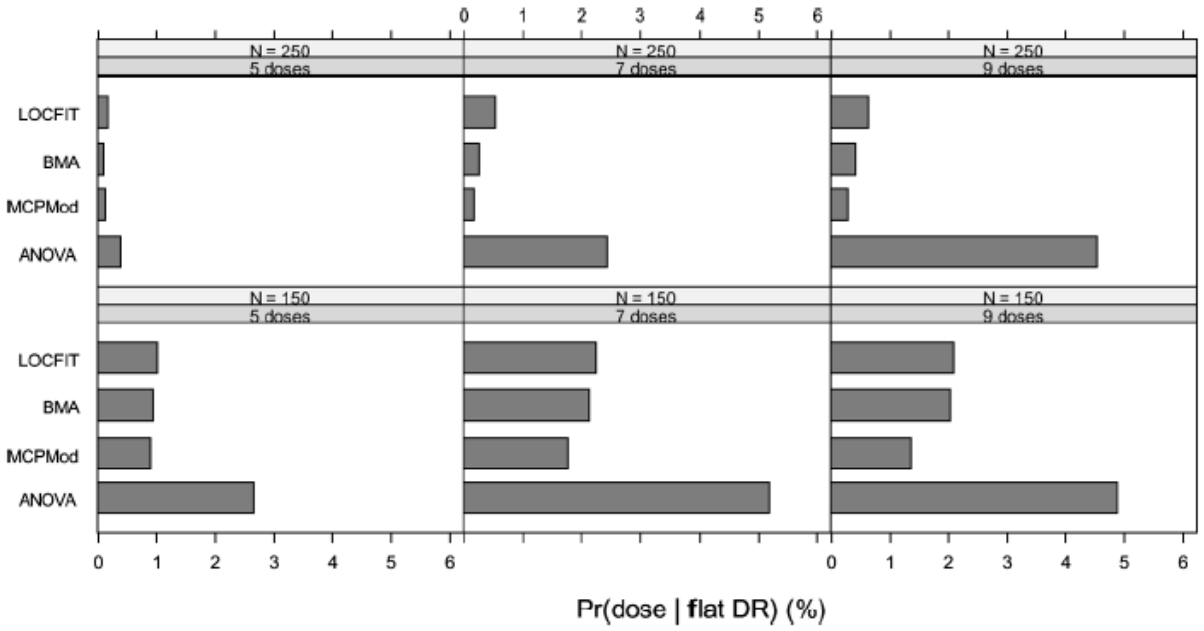

Figure 3-14 Probabilities of identifying clinical relevant dose under flat dose response.

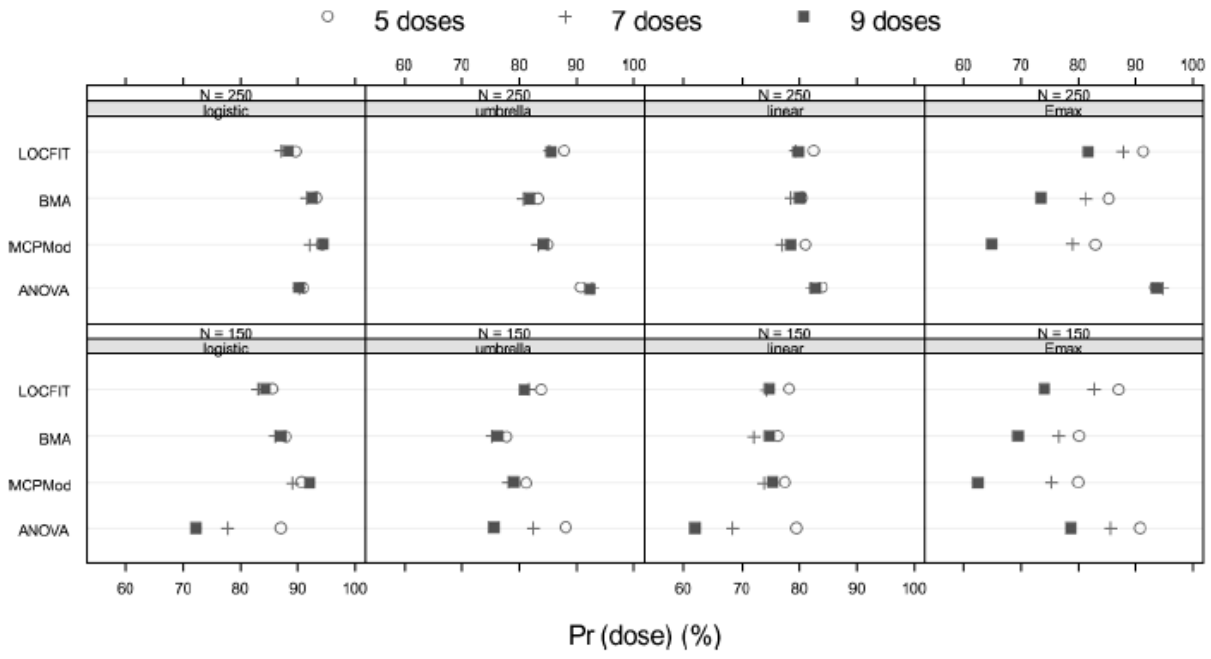

Figure 3-15 Probabilities of identifying clinical relevant dose under active dose response.

(C) Selecting a target dose. The plots of the relative bias and relative absolute error in the target dose estimates are displayed in Figures 3-16 and 3-17, respectively. Both plots indicate

the dependence of the precision of the dose estimate on the dose response profile: the Emax shape leads to considerably more biased and less precise dose estimates, for all methods considered here. Again no clear best method emerges and MCP-Mod is performing quite favorable compared to the other methods.

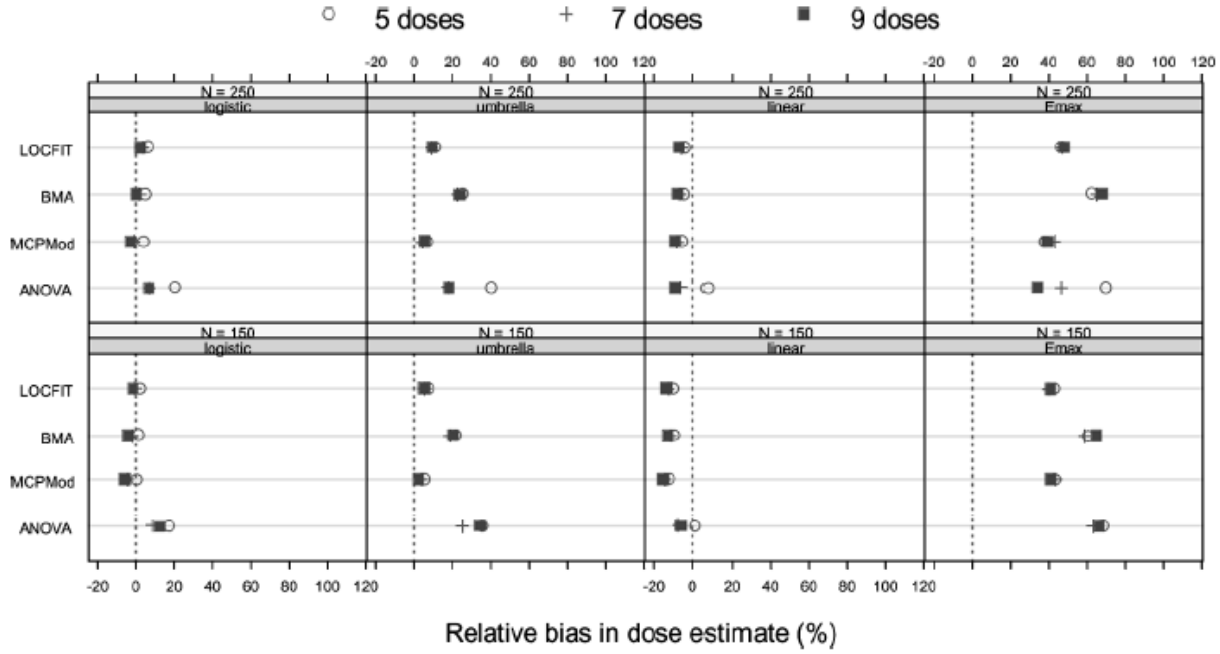

Figure 3-16 Relative bias in target dose estimation.

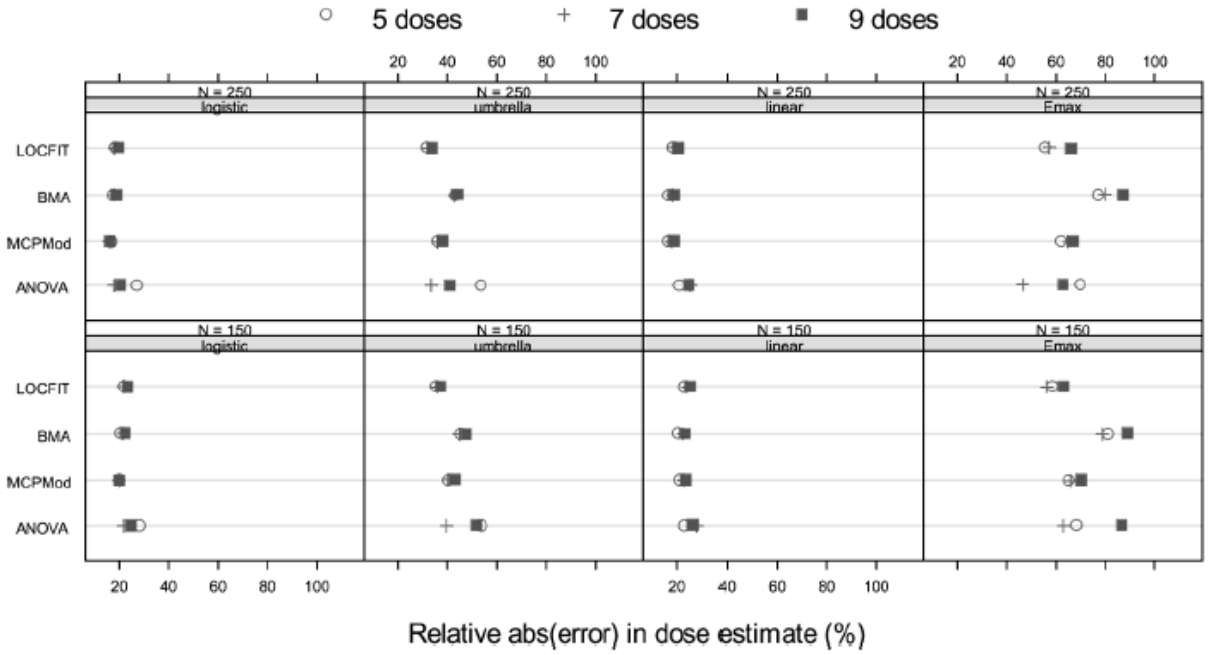

Figure 3-17 Relative absolute error in target dose estimation.

(D) Estimating the dose response. The plots of the average absolute prediction errors (pAPE) relative to the target effect, presented in Figure 3-18, suggest that there are no striking differences among the methods with regard to DR estimation. Also the ANOVA method performs quite acceptable here, which is due to the fact that dose-response modeling was performed for dose-response estimation (in a fashion similar as it would be done for MCP-Mod).

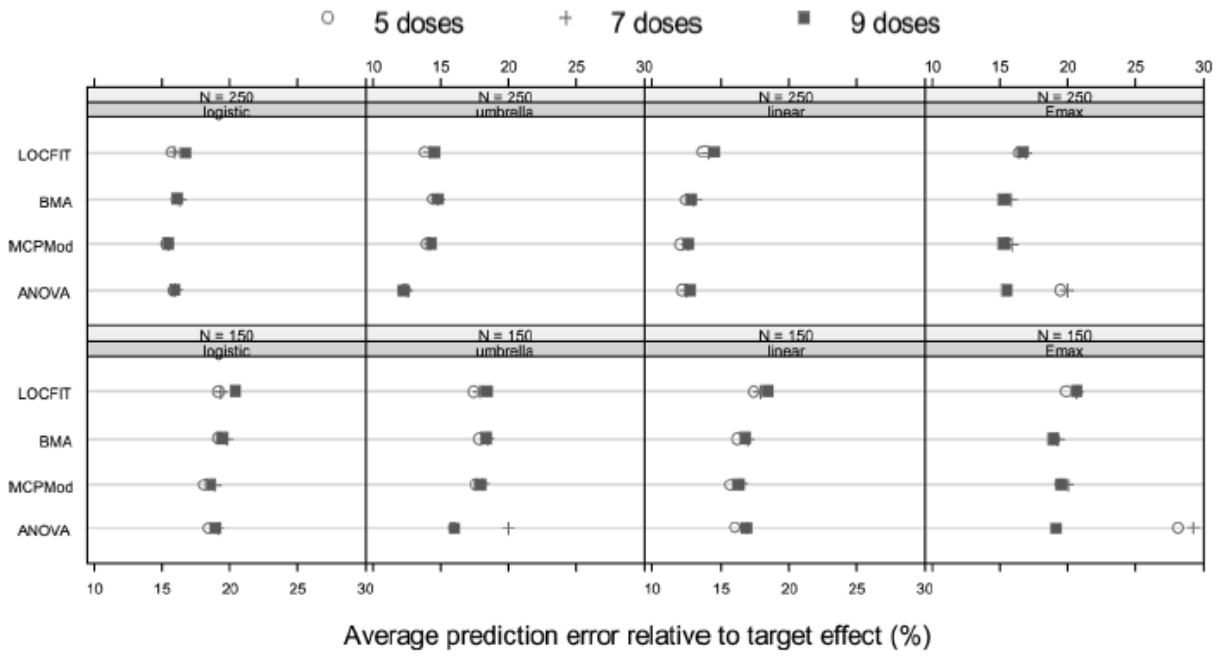

Figure 3-18 Average absolute prediction error relative to target effect.

Overall, we conclude from the PhRMA simulations that MCP-Mod compares favorably in comparison to alternative innovative methods for dose-response signal detection and modeling. In addition, it outperformed the benchmark ANOVA approach in many cases.

4 Conclusions

Current dose finding practices often are not satisfactory: Many dose finding studies are not properly designed (too few doses, incorrect dose-range, sample sizes not large enough, etc.) and the inferences drawn from them not adequately accurate thus leading to major undesirable consequences in drug development (e.g., high failure rate in Phase III, need for label changes after approval, etc.).

Here, we have described a unified strategy for analyzing dose finding studies, including the testing for a dose response signal and the selection of one or more doses to take into further development. The proposed methodology combines the advantages of the multiple comparison and modeling approaches, consisting of a multi-stage procedure. Dose response is tested in the first stage, using multiple comparison methods to identify statistically significant

contrasts corresponding to a set of candidate models. If a dose response signal is established in the first stage, the best model is then used for dose response and target dose estimation. Different model-based dose selection methods, incorporating both statistical significance and clinical relevance, were presented and evaluated via the analysis of real and simulated data.

The MCP part of the proposed method, is seen to maintain the FWER at its nominal level, has power comparable to the standard trend tests under monotone dose-response settings, and is better under non-monotone scenarios, provided the set of candidate models is broad enough. The clear advantage of this new approach, in comparison to more traditional multiple comparison dose finding methods, is its added flexibility in searching for and identifying an adequate dose for future drug development through the use of modeling techniques. Several variants of the proposed methods are possible and have been investigated in the literature.

Compared to traditional ANOVA based pairwise multiple comparisons, MCP-Mod has the advantage of enabling the use of more doses in the design, without requiring a much larger number of patients. In an ANOVA-type approach only the information from the respective dose is used to estimate the dose-response, which means the required sample size depends strongly on the doses studied, when a fixed precision is required at each dose. By using modeling techniques MCP-Mod allows to interpolate information across dose-levels, and the total sample size will not depend strongly on the number of doses studied. The possibility of using more doses will typically result in information-richer dose-finding designs and a better basis for decision making at end of Phase II.

MCP-Mod can be applied to a broad class of parametric models beyond the standard set-up of a normally distributed endpoint in a standard ANOVA model. For example, the response variable might be a count, binary or time-to-event variable instead of being normally distributed. In addition, the final analysis has usually to be adjusted for relevant covariates (e.g. region, age, …), patients measurements are often recorded over time (necessitating the use of longitudinal models), and patients might receive more than one treatment (such as in cross-over or incomplete block designs). Response-adaptive versions of the MCP-Mod approach are also available, which may offer an appropriate compromise to explore wider dose ranges, without making costs prohibitive. On potentially useful application is to conduct seamless Phase IIa / IIb studies, where at an interim analysis proof-of-concept might be declared and, if successful, the information from the first stage is used to finetune the design of the second, dose finding stage.

One implementation of the MCP-Mod approach is available in the `DoseFinding` package in R, which is freely available on CRAN, including the complete source code. This enhances reproducibility and transparency of the underlying calculations. The `DoseFinding` package includes functions for both the design and analysis of dose finding studies using MCP-Mod for general parametric models.

We recommend using MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. It is appropriate to improve current dose finding practices, though it does not provide a remedy for poor design.

5 References

Akacha, M. and Benda, N. (2010) The impact of dropouts on the analysis of dose-finding studies with recurrent event data. Statistics in Medicine 29, 1635-1646

Bartholomew, D.J. (1961) Ordered test in the analysis of variance. Biometrika 48, 325-332.

Bauer, P. and Hackl, P. (1985) The application of Hunters inequality in simultaneous testing Biometrical Journal 27, 25-38.

Benda, N. (2010) Model-based approaches for time-dependent dose finding with repeated binary data. Statistics in Medicine 29, 1096-1106

Bornkamp, B., Bretz, F., Dmitrienko, A., Enas G., Gaydos, B., Hsu, C.H., Koenig, F., Krams, M., Liu, Q., Neuenschwander, B., Parke, T., Pinheiro, J., Roy, A., Sax, R., and Shen, F. (2007) Innovative approaches for designing and analyzing adaptive dose-ranging trials (with discussion). Journal of Biopharmaceutical Statistics 17, 965-995.

Bornkamp, B., Pinheiro, J., and Bretz, F. (2008) MCP-Mod – An R package for the design and analysis of dose finding studies. Journal of Statistical Software 29, 1-23.

Bornkamp, B. and Ickstadt, K. (2009) Bayesian nonparametric estimation of continuous monotone functions with applications to dose-response analysis. Biometrics 65, 198-205.

Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. (2011) Response-adaptive dose-finding under model uncertainty. Annals of Applied Statistics 5(2B), 1611-1631.

Branson, M., Pinheiro, J.C., and Bretz, F. (2003) Searching for an adequate dose: Combining multiple comparisons and modeling techniques in dose-response studies. Novartis Biometrics Technical Report No. 2003-08-20.

Bretz, F., Pinheiro, J., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose response studies. Biometrics 61, 738-748.

Bretz, F., Hsu, J.C., Pinheiro, J.C., and Liu, Y. (2008) Dose Finding - A Challenge in Statistics. Biometrical Journal 50, 480-504.

Bretz, F., Tamhane, A.C., and Pinheiro, J. (2009) Multiple testing in dose response problems. Chapter 3 in: Multiple testing problems in pharmaceutical statistics. Dmitrienko, A., Tamhane, A.C., and Bretz, F. (eds.), Chapman & Hall/CRC Biostatistics Series, Boca Raton, 99-130.

Bretz, F., Hothorn, T., and Westfall, P. (2010) Multiple comparisons using R. CRC Press, Boca Raton.

Buckland, S., Burham, K., and Augustin, N. (1997) Model selection: An integral part of inference. Biometrics 53, 603-618.

Chatfield, C. (1995) Model uncertainty, data mining and statistical inference (with discussion). Journal of the Royal Statistical Society Series A 158, 419-466.

Cross, J., Lee, H., Westelinck, A., Nelson, J., Grudzinkas, C., and Peck, C. (2002) Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999. Pharmacoepidemiology and Drug Safety 11,439-446

Dette, H., Neumeyer, N., and Pilz, K.F. (2005) A note on nonparametric estimation of the effective dose in quantal bioassay. Journal of the American Statistical Association, 100, 503-510.

Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. C. (2008). Optimal designs for dose finding studies. Journal of the American Statistical Association 103, 1225-1237.

Dette, H., Titoff, S., Volgushev, S., and Bretz, F. (2013) Model identification for dose response signal detection. (submitted)

Dragalin, V., Bornkamp, B., Bretz, F., Miller, F., Padmanabhan, S.K., Patel, N., Perevozskaya, I., Pinheiro, J., and Smith, J.R. (2010) A simulation study to compare new adaptive dose-ranging designs. Statistics in Biopharmaceutical Research 2, 487-512.

Draper, D. (1995) Assessment and propagation of model uncertainty. Journal of the Royal Statistical Society, Series B 57, 45-97.

Hastie, T. and Loader, C. (1993) Local Regression: Automatic Kernel Carpentry. Statistical Science 8, 120-129.

Heerdink, E.R., Urquhart, J., and Leufkens, H.G. (2002) Changes in prescribed dose after market introduction. . Pharmacoepidemiology and Drug Safety 11, 447-453

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999) Bayesian model averaging: a tutorial (with discussion). Statistical Science 14, 382-417.

ICH-E4 (1994). ICH Harmonized Tripartite Guideline. Topic E4: Dose-response information to support drug registration. Available at www.emea.eu.int

Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose response curves and the assessment of synergism. Biometrics, 46:1071-1085.

Klingenberg, B. (2009) Proof of concept and dose estimation with binary responses under model uncertainty. Statistics in Medicine 28, 274-292.

Mercier, F., Ohlssen, D., Bornkamp, B., Rochotte, E., Li, D., Wallstroem, E., and Pinheiro, J. (2013) Characterization of dose-response for count data using generalized MCP-Mod approach in an adaptive dose-ranging trial (submitted)

Miller, F. (2010) Adaptive dose-finding: Proof of concept with type I error control. Biometrical Journal 52, 577-589.

Mukhopadhyay, S. (2000) Bayesian nonparametric inference on the dose level with specified response rate. Biometrics 56, 220-226.

Pinheiro, J., Bretz, F., and Branson, M. (2006a). Analysis of dose response studies: Modeling approaches. pp. 146–171 in Ting, N. (ed.) Dose finding in drug development. Springer, New York.

Pinheiro, J., Bornkamp, B., and Bretz, F. (2006b). Design and analysis of dose finding studies combining multiple comparisons and modeling procedures. Journal of Biopharmaceutical Statistics 16, 639-656.

Pinheiro, J., Sax, F., Antonijevic, Z., Bornkamp, B., Bretz, F., Chuang-Stein, C., Dragalin, V., Fardipour, P., Gallo, P., Gillespie, W., Hsu, C.H., Miller, F., Padmanabhan, S.K., Patel, N.,

Perevozskaya, I., Roy, A., Sanil, A., and Smith, J.R. (2010) Adaptive and model-based dose-ranging trials: Quantitative evaluation and recommendations. White Paper of the PhRMA Working Group on "Adaptive Dose-Ranging Studies" (with discussion). Statistics in Biopharmaceutical Research 2, 435-454.

Pinheiro, J., Bornkamp, B., Glimm, E., and Bretz F (2013) Multiple comparisons and modeling for dose finding using general parametric models (submitted)

Prinz, F., Schlange, T., and Asadullah, K. (2011) Believe it or not: how much can we rely on published data on potential drug targets? Nature Reviews Drug Discovery 10, 712-713

Ruberg, S.J. (1989) Contrasts for identifying the minimum effective dose. Journal of the American Statistical Association 84, 816-823.

Ruberg, S.J. (1995a) Dose response studies I. Some design considerations. Journal of Biopharmaceutical Statistics 5, 1-14.

Ruberg, S.J. (1995b) Dose response studies II. Analysis and interpretation. Journal of Biopharmaceutical Statistics 5, 15-42.

Stewart, W. H. and Ruberg, S. J. (2000) Detecting dose response with contrasts. Statistics in Medicine 19, 913-921.

Tanaka, Y. (2011) An Adaptive Two-Stage Dose-Response Design Method for Establishing Proof of Concept in Drug Development. PhD thesis, University of Pittsburgh.

Tanaka, Y. and Sampson, A. (2013) An adaptive two-stage dose-response design method for establishing proof-of-concept. (submitted)

Tao, A. (2010) A dose finding method in joint modeling of efficacy and safety endpoints in Phase II studies - An extension of the MCP-MOD method. PhD thesis, School of Public Health, University of Medicine and Dentistry of New Jersey

Thomas, N. (2006). Hypothesis testing and Bayesian estimation using a sigmoid Emax model applied to sparse dose response designs. Journal of Biopharmaceutical Statistics 16, 657-677.

Tukey, J. W., Ciminera, J. L., and Heyse, J. F. (1985). Testing the statistical certainty of a response to increasing doses of a drug. Biometrics 41, 295-301.

Wakana, A., Yoshimura, I., and Hamada, C. (2007). A method for therapeutic dose selection in a phase II clinical trial using contrast statistics. Statistics in Medicine 26, 498-511.

Wouter, W. (2012) The power and type I error of statistical analysis approaches for longitudinal toxicological data. M.Sc. thesis, University of Gent.

Yuan, Y. and Yin, G. (2011). Dose-response curve estimation: A semiparametric mixture approach. Biometrics, 67:1543-1554.

6 Appendix

Model-name	μ(d, **θ**)	(*)
linear	$E_0 + \delta d$	
linlog	$E_0 + \delta \log(d + c)$	
quadratic	$E_0 + \beta_1 d + \beta_2 d^2$	δ
emax	$E_0 + E_{max} d/(ED_{50} + d)$	ED_{50}
logistic	$E_0 + E_{max}/\{1 + \exp[(ED_{50} - d)/\delta]\}$	$(ED_{50}, \delta)'$
exponential	$E_0 + E_1(\exp(d/\delta) - 1)$	δ
sigEmax	$E_0 + E_{max} d^h/(ED_{50}^h + d^h)$	$(ED_{50}, h)'$
betaMod	$E_0 + E_{max} B(\delta_1, \delta_2)(d/D)^{\delta_1}(1 - d/D)^{\delta_2}$	$(\delta_1, \delta_2)'$

Table 6-1 **Dose-response models implemented in the DoseFinding package. Column (*) lists for each model the parameters that determine the shape of the model. 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** $\delta = \frac{\beta_2}{|\beta_1|}$**.**