Comparison of different approaches for dose response analysis

Saswati Saha | Werner Brannath

Abstract
Characterizing an appropriate dose-response relationship and identifying the right dose in a clinical trial are two main goals of early drug-development. MCP-Mod is one of the pioneer approaches developed within the last 10 years that combines the modeling techniques with multiple comparison procedures to address the above goals in clinical drug development. The MCP-Mod approach begins with a set of potential dose-response models, tests for a significant dose-response effect (proof of concept, PoC) using multiple linear contrasts tests and selects the “best” model among those with a significant contrast test. A disadvantage of the method is that the parameter values of the candidate models need to be fixed a priori for the contrasts tests. This may lead to a loss in power and unreliable model selection. For this reason, several variations of the MCP-Mod approach and a hierarchical model selection approach have been suggested where the parameter values need not be fixed in the proof of concept testing step and can be estimated after the model selection step. This paper provides a numerical comparison of the different MCP-Mod variants and the hierarchical model selection approach with regard to their ability of detecting the dose-response trend, their potential to select the correct model and their accuracy in estimating the dose response shape and minimum effective dose. Additionally, as one of the approaches is based on two-sided model comparisons only, we make it more consistent with the common goals of a PoC study, by extending it to one-sided comparisons between the constant and alternative candidate models in the proof of concept step.

KEYWORDS
dose finding studies, log likelihood models, MCP-Mod, model selection

1 | INTRODUCTION

A major concern in drug-development is characterization of the dose-response curve. Phase II dose finding studies often aim to detect a dose-response trend and simultaneously estimate the dose-response relationship. The goal of such Proof of Concept (PoC) studies is to determine whether to continue with the drug-development process and which dose should be selected for later stages.

Dose-finding studies were analyzed historically by two approaches: (i) Multiple comparisons procedure (MCP) and (ii) Modeling Techniques. Under the MCP approach, each active dose is compared to placebo and PoC is established using multiple testing methods (Hothorn, Neuhäuser, & Koch, 1997; Stewart & Ruberg, 2000). Once the PoC is established, the minimum active dose with a statistically significant outcome and clinically relevant effect compared to the placebo is selected as the minimum effective dose (MED) (Ruberg, 1995; Tamhane, Hochberg, & Dunnett, 1996; Tamhane & Logan, 2002). Such procedures are relatively robust to the underlying dose-response relationship and are more effective when there are few active doses in the study. But they are not appropriate for characterizing the dose-response relationship. Dose-response inference is confined to the
distinct doses administered in a given trial. The modeling approach on the other hand establishes PoC by comparing the flat dose-response model against the fitted dose-response model (Morgan, 1992; Pinheiro, Bretz, & Branson, 2006b). If the PoC is established, the underlying regression model is utilized for further inference on dose-response profiles. Figure 1 demonstrates the dose response models that are considered most relevant in the context of dose finding studies. The reliability of the model-based approach depends on the correct specification of the dose-response regression model that is not known at the beginning of the trial.

Motivated by the shortcomings of the historical approaches, Bretz, Pinheiro, and Branson (2005) proposed an integrated approach namely the MCP-Mod (Multiple Comparison Procedures with Modeling) approach. This approach combines the multiple comparisons and modeling techniques and comprises two steps: (i) They start with a set of potential parametric dose-response models and conduct multiple linear contrast tests for testing a dose-related effect (PoC). (ii) Once the PoC is established, the best model (if any) is obtained from the set of models with significant contrasts either by model selection criteria or by model averaging. (iii) Upon selecting the best model, the adequate doses are estimated based on regression techniques. An implementation of MCP-Mod and its application is illustrated in Pinheiro, Bornkamp, and Bretz (2006a). MCP-Mod is recently qualified by the European Medicines Agency as an efficient statistical methodology for model-based design and analysis of phase II dose-finding studies under model uncertainty (EMA, 2014). However, a major deficiency of the MCP-Mod approach is that it makes prior assumptions about the values of the model parameters used in the candidate set. These unknown parameters, if wrongly specified, may lead to inefficient inferences in the dose-finding step.

In this article, we will discuss three dose-response profile estimation methods which address the shortcomings of MCP-Mod method. The first approach suggested by Dette, Titoff, Volgushev, and Bretz (2015) is a variation of the MCP-Mod approach that uses likelihood ratio tests instead of prespecified contrasts tests used in the original MCP-Mod. Thus this method does not need prior knowledge of the parameters of the candidate models. This is referred to as multiple comparison likelihood ratio asymptotic (MCLRa) method in the rest of the paper. The second approach proposed by Baayen, Hougaard, and Pipper (2015) compares and evaluates a candidate set of nested, plausible monotone dose-response shapes against the constant model of no effect and PoC is established if the constant model is rejected. The candidate models are also compared among each other via sequential testing procedure to come up with one final model which best fits the data under the given set up. This is referred to as multiple hierarchical modeling and comparison (MHMC) method in the rest of the paper. The MCLRa and MHMC method both use the asymptotic distribution of log-likelihood ratio statistics for their PoC testing. The third approach proposed by Gutjahr and Bornkamp (2017) considers the exact distribution of the LR test for the PoC testing step. Here, the author uses LR test statistics for testing the null hypothesis that there is no dose-response trend against the composite alternative that one of the candidate
dose-response shapes is true. This is referred to as multiple comparison likelihood ratio exact (MCLRe) method in the rest of the paper. The MCLRa and MCLRe methods possess the property of testing the evidence of dose-response trend and estimation of target dose similar to MCP-Mod. Thus they are referred to as MCP-Mod variants in the rest of the paper. The three methods are compared in this paper amongst each other and with the original MCP-Mod approach with respect to their capability to: (i) detect a nonzero dose-response trend, (ii) identify the correct model, (iii) estimate the dose-response shape, and (iv) estimate the target dose like the minimum effective dose (MED).

The rest of the paper is organized as follows: The dose-finding methods compared in this paper are reviewed in Section 2. Section 3.1 describes the design of the simulation studies and Section 3.2 describes the performance metrics used in the comparison of the different MCP-Mod variants. The results of the analyses are presented in Section 3.3. The paper concludes with a discussion in Section 4.

2 | DOSE RESPONSE ANALYSIS METHODS

2.1 | Basic modeling framework

Consider a random vector $Y$ containing the clinical measurement of interest for each of the $n$ patients. The response $Y$ is observed for $k + 1$ parallel dose groups and is assumed to follow a certain dose-response curve with normally distributed errors:

$$Y_{ij} = \mu_l(d_i, \theta_l) + \epsilon_{ij} = \alpha + \beta x_{\gamma_l}(d_i) + \epsilon_{ij},$$

$$\epsilon_{ij} \sim N(0, \sigma^2), \quad i = 0, \ldots, k, \quad j = 0, \ldots, n_i,$$

where $\mu_l(d_i, \theta_l)$ denotes the regression function under model family $l$ and dose $d_i$, $d_0$ denotes the placebo dose and $d_1, d_2, d_3, \ldots, d_k$ are the active doses investigated in the parallel group design and $\theta_l = \{\alpha, \beta, \gamma_l; \alpha, \beta \in \mathbb{R}, \gamma_l \in \Gamma_l \subseteq \mathbb{R}^{p-2}\}$ refers to the vector of model parameters. $Y_{ij}$ refers to the response of the $j^{th}$ patient in the dose group $i$ where $n_i$ is the number of patients in dose group $i$. Table 1 shows examples of linear or nonlinear transformations of the dose variable $d$ that can be considered in $x_{\gamma_l}(d)$.

2.2 | Choice of candidate set

The methods discussed in this paper consider a set of parametric model families known as the candidate set of dose-response models to capture the model uncertainty more precisely. The candidate set is denoted by $\mathcal{M} = \{M_l : l = 1, 2, \ldots, m\}$ and the mean response vector for the $l^{th}$ model family is denoted by $\bar{\mu}_l = (\mu_{l0}, \ldots, \mu_{lk})$ where $\mu_{li}$ is the mean response corresponding to dose group $i$ if $M_l$ is the correct model family. Table 1 shows few examples of parametric models in the candidate set used in the MCLRa and MCLRe method and Table 2 shows an example of a set of nested, monotone candidate models used in the MHMC method. Three different candidate sets are considered in this paper for evaluating the different methods; see Table 3. The motivation for choosing each candidate set is explained below:

**Candidate Set 1:**

In this comparison models with three parameters are only considered in the candidate set of MCP-Mod, MCLRa, and MCLRe method because PoC evaluation of the MCLRe method is time consuming for models with more parameters. The MHMC method is using the candidate set shown in Table 2. This candidate set was also used for the MHMC method in the original

| Model               | $\mu_l(\alpha, \beta, \gamma_l, d)$ | $x_{\gamma_l}(d)$ |
|---------------------|-------------------------------------|-------------------|
| Linear              | $\alpha + \beta d$                  | $d$               |
| Linear in log-dose  | $\alpha + \beta \log(d + c)$       | $\log(d + c)$     |
| Exponential         | $\alpha + \beta \left[\exp\left(\frac{d}{\delta}\right) - 1\right]$ | $\left[\exp\left(\frac{d}{\delta}\right) - 1\right]$ |
| Quadratic           | $\alpha + \beta d + \frac{\beta_2}{\delta} d^2$ | $d + \frac{\beta_2}{\delta} d^2$ for $\beta_2 < 0$ |
| Sigmoid Emax        | $\alpha + \beta \frac{d^3}{d + \delta^3}$ | $\frac{d^3}{d + \delta^3}$ |
**Table 2** Candidate set for MHMC method

| Model | Function |
|-------|----------|
| $M_0$ | No effect $\alpha$ |
| $M_1$ | Linear $\alpha + \beta d$ |
| $M_2$ | Power $\alpha + \beta d^{\gamma}$ |
| $M_3$ | Four parameter logistic $\alpha + \beta \left( \frac{d}{d + E_{D_{50}}} \right)^{\gamma}$ |
| $M_4$ | Unrestricted model $\mu_l(\alpha, \beta, \gamma : d_i) = \mu_i$ |

**Table 3** Candidate set for different methods

| Candidate set | MCP-Mod | MCLRa | MCLRe | MHMC |
|---------------|---------|-------|-------|------|
| Candidate set 1 | $d$ | $d$ | $d$ | $d$ |
| | $\log(d + 0.2)$ | $\log(d + c)$ | $\log(d + c)$ | $d^\gamma$ |
| | $\frac{d}{0.2 + d}$ | $\frac{d}{E_{D_{50}} + d}$ | $\frac{d}{E_{D_{50}} + d}$ | $\frac{d}{(d^{\gamma - 1} + 1)}$ |
| | $\exp\left( \frac{d}{0.28} - 1 \right)$ | $\exp\left( \frac{d}{\gamma} - 1 \right)$ | $\exp\left( \frac{d}{\gamma} - 1 \right)$ | $\exp\left( \frac{d}{\gamma} - 1 \right)$ |
| Candidate set 2 | $d$ | $d$ | $d^\gamma$ | $d^\gamma$ |
| | $\log(d + 0.2)$ | $\log(d + c)$ | $\log(d + c)$ | $d^\gamma$ |
| | $\frac{d}{0.2 + d}$ | $\frac{d}{E_{D_{50}} + d}$ | $\frac{d}{E_{D_{50}} + d}$ | $\frac{d}{(d^{\gamma - 1} + 1)}$ |
| | $\exp\left( \frac{d}{0.28} - 1 \right)$ | $\exp\left( \frac{d}{\gamma} - 1 \right)$ | $\exp\left( \frac{d}{\gamma} - 1 \right)$ | $\exp\left( \frac{d}{\gamma} - 1 \right)$ |
| Candidate set 3 | $d$ | $d$ | $d$ | $d$ |
| | $\frac{d}{0.2 + d}$ | $\frac{d}{E_{D_{50}} + d}$ | $\frac{d}{E_{D_{50}} + d}$ | $\frac{d}{E_{D_{50}} + d}$ |
| | $\frac{d}{0.4 + d^2}$ | $\frac{d}{E_{D_{50}} + d^2}$ | $\frac{d}{E_{D_{50}} + d^2}$ | $\frac{d}{E_{D_{50}} + d^2}$ |

**Table 4** Computation time (in seconds) of different methods for running 1 simulation with candidate set 1

| Sample size per dose | MCP-Mod | MCLRa | MCLRe | MHMC |
|----------------------|---------|-------|-------|------|
| 10                   | 0.05    | 0.03  | 0.24  | 1.02 |
| 25                   | 0.05    | 0.02  | 0.80  | 1.20 |
| 50                   | 0.10    | 0.03  | 2.87  | 0.78 |
| 75                   | 0.07    | 0.03  | 4.73  | 0.87 |
| 100                  | 0.06    | 0.03  | 7.94  | 1.38 |

article (Baayen et al., 2015). So candidate set 1 primarily focuses on the performance of MCLRe method as compared to the other methods.

**Candidate Set 2:**

In the second comparison, complex nonlinear models with more parameters like sigmoidal Emax (sigEmax) model are also considered in the candidate set of MCP-Mod and MCLRa method. The objective is to check if there is any further information gained on widening the candidate set of the above methods. For the MHMC method, the candidate set remain the same as mentioned in Table 2. MCLRe is no longer considered in this comparison because the performance of MCLRe is similar to MCLRa. Additionally, the MCLRe method is more time consuming compared to the other methods (see Table 4).

**Candidate Set 3:**

In the third candidate set (see Table 3) the power model in Table 2 is replaced by the emax model for the MHMC method. By this, the candidate set of MHMC and MCLRa are becoming quite similar, as the four parameter logistic (4PL) model (see Table 2) and sigmoidal Emax model lead to the same class of dose-response models. Moreover the nested nature of the candidate set is preserved. This was already mentioned in Baayen et al. (2015), but not investigated. The linlog model is excluded from the candidate set of the MCPMod and MCLRa method since emax and linlog are very close to each other and the emax dose-response shape is more common in Phase II dose-response studies.
2.3 Methods

Method based on LR contrast test (MCLRa)

The method suggested by Dette, Titoff, and Volgushef is a variation of the MCP-Mod approach as already mentioned in the introduction. They use likelihood ratio test instead of the $t$ test used in the PoC testing step. This is addressed as LR contrast test in their paper (Dette et al., 2015). Given that the model specifications in the earlier section are true and the objective is to choose a model among the candidate models in $\mathcal{M}$, the hypothesis tested to detect the dose response signal is:

$$H_{0,l} : e_l^T \mu_l = 0 \quad \text{against} \quad H_{1,l} : e_l^T \mu_l > 0$$

for a given contrast $e_l = (c_{l,1}, ..., c_{l,p})^T \in \mathbb{R}^p$.

Let $\| \|_2$ denote the Euclidean norm. The LR test statistic for testing the above hypothesis is given by:

$$LR_{n,l} = -2 \log \frac{T_{n,H_0}^l}{T_{n,H_1}^l} \quad l = 1, 2, ..., m$$

where

$$T_{n,H_0}^l = \sup \left\{ \frac{1}{\sigma^n} \exp \left( -\frac{1}{2\sigma^2} \| Y - \mu_l(\theta_l) \|_2^2 \right) \left| \sigma > 0; \theta_l \in \mathbb{R}^{d_l}; e_l^T \mu_l = 0 \right. \right\},$$

$$T_{n,H_1}^l = \sup \left\{ \frac{1}{\sigma^n} \exp \left( -\frac{1}{2\sigma^2} \| Y - \mu_l(\theta_l) \|_2^2 \right) \left| \sigma > 0; \theta_l \in \mathbb{R}^{d_l}; e_l^T \mu_l > 0 \right. \right\}.$$

The LR test rejects the null hypothesis of no dose response for large values of the test statistics $LR_{n,l}$. In order to conclude in favor of a statistically significant dose response signal, $m$ individual LR test statistics are combined into a maximum statistics:

$$LR_{n,max} = \max_{l=1}^m LR_{n,l}.$$ (2)

In classical likelihood theory, the statistic $LR_{n,l}$ usually converges weakly to a $\chi^2$ distribution provided the parameters characterizing the null distribution are unique. But it is well known that classical asymptotic theory are not applicable when parameters are unidentifiable under null hypothesis of the above set up (Andrews & Ploberger, 1994; Chernoff, 1954; Davies, 1987; Lindsay, 1995). The authors have applied nonstandard asymptotic results (Liu & Shao, 2003) to address the nonidentifiability in the LR contrast tests. They have shown that the asymptotic distribution of the likelihood ratio test statistic (1) corresponding to each model can be approximated by a functional of a stochastic process. The covariance between the likelihood ratio of the two models can be approximated by the covariance between two stochastic process that has a closed kernel equation. The derivations are described in detail in the original manuscript (Dette et al., 2015). The resulting limiting quantile of the max LR statistic can be calculated as:

$$LR_{max} = \max_{l=1}^m LR_l \quad \text{where} \quad LR_l \quad \text{is the limiting distribution of} \quad LR_{n,l}.$$ (3)

Following the PoC testing step, the model selection in this method can be done based on $P$-value or some model selection criteria like Akaike Information Criteria (AIC). Model selection based on $P$-values is highly criticized in literature (Baayen et al., 2015) since models are not compared with each other when selection is done based on the smallest $P$-value, so model selection based on the AIC criteria is suggested in this paper. Thus if any significant result is obtained at the PoC testing step then the “best” model is selected from the set of significant models and this model is then deployed to obtain the target dose(s) (MED in this case) using inverse regression techniques.

Method based on LR exact test (MCLRRe)

As discussed in the earlier section, the parametric models permissible in the candidate set of LR contrast test and LR exact test method are similar. The null hypothesis tested here is:

$$H_0 : \beta = 0 \Leftrightarrow H_0 : \mu \in \{ a 1_n : a \in \mathbb{R} \},$$
and the one-sided alternatives \( \mu \in H_1, ..., \mu \in H_m \), where, for \( i = 1, ..., m \),
\[
H_i = \{ \mu_i(\theta_i) = \mu_i(a, \beta, \gamma) : a \in \mathbb{R}, \beta \in \mathbb{R}_+, \gamma \in \Gamma_i \},
\]
and the multiple alternative \( \mu \in H_1 \), where \( H_I = \bigcup_{i \in I} H_i \), for \( I \subseteq \{1, ..., m\} \).

The goal is to derive the likelihood-ratio test for \( H_0 \) against \( H_I \). The above test can be simplified as \( H_{0, I} : y \sim \mathcal{N}(\mu_1, \sigma^2 I_n) \) against \( H_{I, I} : y \sim \mathcal{N}(\mu_1(\alpha, \beta, \gamma), \sigma^2 I_n) \). The authors showed that LR test statistic used in the above testing simplifies to:
\[
S(R_i) = (1 - 1(R_i > 0)R_i^2)^{n/2} \text{ where } R_i = \sup_{\gamma_i \in \Gamma_i} \tilde{x}_{i,d}^T \tilde{y} \quad \text{and} \quad \tilde{x}_{i,d} = \frac{B_x y_{i,d}}{\|B_x y_{i,d}\|}, \quad \tilde{y} = \frac{B_y}{\|B_y\|}
\]
where \( \gamma_i \) is the nonlinear parameter defined earlier and \( \Gamma_i \) is the parameter space for \( \gamma_i \). The rows of the \((n-1) \times n\) matrix \( B \) form an orthonormal basis for the linear subspace \( L = \{ a \in \mathbb{R}^n : a'1_n = 0 \} \). The LR test rejects if \( S \) is small enough or equivalently since \( S \) is non-increasing, if \( R \) is large enough. The authors started with the computation of \( P \)-value for single one-sided alternative with single dose response model and extended their methods to multiple dose response models. For multiple alternatives with multiple dose response models the test statistics used is:
\[
S(R) = \min_{i \in I = \{1, ..., m\}} S(R_i) \Leftrightarrow R = \max_{i \in I} R_i = \max_{i \in I} \sup_{\gamma_i \in \Gamma_i} \tilde{x}_{i,d}^T \tilde{y}
\]
The \( P \)-value of the LR test is given by \( P_0(R > r) \) where \( r \) is the observed value of \( R \) and \( P_0() \) is the probability calculated under the null hypothesis \( H_0 \). The authors have shown that the distribution of the test statistic for PoC can be determined using results from differential geometry. Further details are given in the original paper (Gutjahr & Bornkamp, 2017).

To the best of our knowledge, there exists only few literature on exact LR test involving nonlinear models. The approach used here can be used in applications beyond dose-response modeling. For instance, trend tests involving nonlinear models arise in many areas of applied sciences like change point analysis. One can apply similar approach like the MCLRe method to compute the distribution of the test statistic in the above scenario. This approach also provides a way to evaluate the power of the proposed test under alternative hypothesis, thereby enabling sample size calculations that are essential for clinical trials. The model selection and dose-response inference in this method is similar to the MCLRa approach.

**Method based on hierarchical model testing (MHMC)**

As shown in the earlier section the choice of candidate set for this method is not so flexible as the earlier approaches. This is because monotonicity and nestedness are needed for the candidate set used in this method. Table 2 gives an example of a candidate set of models that can be used in MHMC method. However, the candidate set is not fixed here. Any set of nested, monotone models can be considered as candidate set. The testing method sequentially starts with comparing the constant model to the other complex models in the candidate set and proceeds with the testing and comparison of more complex models in the subsequent testing steps:
\[
H_{s,0} : f(d, \theta) = M_s \\
H_{s,a} : f(d, \theta) = M_r \quad \text{for any} \quad r \in s + 1, ..., m
\]
starting with \( s = 0 \) and going up to \( s = m - 1 \), where models are ordered according to their complexity.

The fixed sequence test mentioned above evaluates the PoC at the first step and if a significant effect is found then the process proceeds selecting the final model from the candidate set. The test statistics used are:
\[
T_s = \max_{s+1 \leq r \leq m} \frac{2(L_r - L_s)}{(p_r - p_s)}
\]
with \( p_r \) and \( p_s \) the number of parameters in model \( r \) and \( s \) respectively and \( L_r \) and \( L_s \) are the log likelihood of the model \( r \) and \( s \) respectively, evaluated at the maximum likelihood estimates of the model parameters of the respective model.

Note that, to account for over fitting, the above method penalizes for large differences in the number of model parameters by the choice of the denominator in the test statistics. The virtue of this method lies in the fact that both testing and model selection is done in the same statistical framework and model parameters are not specified at the beginning. Similar to the MCLRa approach, the author have also used nonstandard asymptotic results to obtain the asymptotic distribution of the log-likelihood ratio tests under nonidentifiability. They have presented in their original paper (Baayen et al., 2015) a characterization of the distribution of the log-likelihood ratio test statistic for two-sided PoC test under \( H_0 \), when the requirement \( \beta = 0 \) renders the other nonlinear
### Table 5 Data generating dose-response shapes

| Model       | \( \mu_l(\alpha, \beta, \gamma, d) \) |
|-------------|--------------------------------------|
| constant    | 0.2                                  |
| linear      | 0.2 + 0.6d                           |
| linlog      | 0.74 + 0.33log(d + 0.2)              |
| exponential | 0.183 + 0.017exp\left(\frac{d}{0.23}\right) |
| emax        | 0.2 + 0.7\frac{d}{0.23+d}           |
| emax2       | 0.1 + 0.5\frac{d}{0.03+2d}          |
| sigmoidal Emax | 0.2 + 0.6\frac{d}{0.03+2d}         |
| power       | 0.2 + 0.6d^{0.5}                     |
| truncated logistic | 0.2 + \frac{0.6d}{1+exp(100.8−d))} |
| quadratic   | 0.2 + 2.049d - 1.749d²               |

The authors have also considered an unrestricted model \( M_4 \) shown in Table 2. They have shown in their analyses that inclusion of the unrestricted model protects the method against model misspecification and increases the power to detect the dose-response trend, if the true dose-response shape (a 4PL model or any model nested in 4PL model for instance) is not covered by the candidate set. However, we could not find sufficient evidence in support of the above claim in our simulations. Further, a non-monotone dose-response shape is not very common in clinical trials so it might not be necessary to safeguard the method against a non-monotone shape. However, the unrestricted model \( M_4 \) might be useful to increase robustness with regards to the assumed candidate models.

### 3 SIMULATION STUDIES

In this section, we present the simulation study for the comparison of the MCP-Mod variants and the MHMC method introduced in Section 2. We compare them with respect to their power to detect a nonflat dose-response shape (PoC). We also consider their ability to select the right model from the candidate set and study the model selection behavior when the true model belongs to the candidate set. Since the dose-response shape may be approximated sufficiently well even when the wrong model has been selected, we also consider the cumulative absolute deviation between the estimated and true model shape. We compare this prediction error to the error that we could have achieved if we would have known the right model family (with unknown parameter values) from the very beginning. Moreover, we also investigate how well a dose-response shape can be approximated if the true shape is not a member of the candidate set. For instance, exponential model is not included in the Candidate Set 3 (see Table 3) but we simulated the data from an exponential model (see Table 5) and compared the closeness between the predicted model and the correct model under the different methods.

Section 3.1 describes the design of the simulation study with the assumptions and the scenarios considered in this article; Section 3.2 describes the performance metrics used to evaluate their performance. Graphical representation of the statistical performance of each method based on the simulation results are shown at the end of the section.

### 3.1 Design of the study

#### 3.1.1 Study design

The study design is based on the design suggested in Bretz et al. (2005). The simulation scenarios are based on a randomized double-blind parallel group trial with a total of \( n \) patients being allocated to either placebo or one of the four active doses coded as 0.05, 0.20, 0.60, and 1. The methods are evaluated across five sample sizes: \( n_j = 10, 25, 50, 75, 100 \) per dose-group. The response variable is assumed to be normally distributed as specified in Section 2.1. A standard deviation of 1.478 is used in the simulation for assessing the power to detect the trend. This standard deviation gives a power of 80% at a sample size of 75 for a pairwise test between two doses. For the purpose of evaluating the model selection and dose-selection performance of the different methods, a response SD of \( \sigma = 0.65 \) was used. Two different standard deviation were considered to make the...
simulation study consistent with the original MCP-Mod article (Bretz et al., 2005) where similar assumptions were adopted while simulating the scenarios. For each dose-response shape and sample size combination, shown in the subsequent section, 10,000 simulations are considered.

### 3.1.2 Data generating shapes

Data are simulated from nine nonlinear models shown in Figure 1 and Table 5. Note that the second emax model, emax2, is not a desirable dose-response model in a typical clinical trial. It is not desirable from a clinical point of view because it does not reach the anticipated efficacy like the other models. However, the methods should identify such extreme situations as well. So it is included in our study. Similarly the truncated logistic model (tlog) and the quadratic model is included to check how the methods are performing for models not included in the candidate set. Note that the first model is monotonic while the second one is not.

All the shapes except the second emax shape (emax2) and the constant shape share the property that the response is 0.2 at the placebo dose \( d = 0 \) and the maximum response is attained at 0.8.

### 3.2 Performance evaluation of the methods

The main goal of the above methods is to attain PoC and identify the dose-response profile efficiently. To quantify the performance of the different methods in achieving the above, the following performance metrics are analyzed in the simulation study:

1. **Detecting Dose Response Trend (DRT):** The probability of detecting a positive dose response trend \( P(DRT) \) is estimated as the proportion of simulated trials in which the decision rule concludes in favor of dose-response activity. Under the constant model, \( P(DRT) \) gives the type 1 error rate. We always consider the one sided PoC tests at 5% significance level.

2. **Estimating the dose-response profile:** If the PoC is established, two measures are computed to evaluate the accuracy in estimating the dose-response shape: the model selection probability and the cumulative absolute prediction error. The model selection probability is estimated as the proportion of simulated trials in which the correct model is selected after PoC is established. The methods are compared with respect to this model selection probability.

   If the wrong model has been selected from the candidate set then the method can either fail to provide a good estimate of the dose response shape or can sufficiently well approximate the correct model. Therefore we suggest to investigate the precision of the model shape estimation. As a performance measure we use the cumulative absolute prediction error (\( CAPE \)). It is defined as:

\[
CAPE = \int_0^1 |μ(x, \hat{θ}) - μ_0(x, θ_0)| \, dx,
\]

where \( μ_0(x, θ_0) \) denotes the mean response at dose \( x \) under the actual dose-response model (\( μ_0 \)) and \( μ(x, \hat{θ}) \) denotes the predicted response at dose \( x \) based on the dose-response model (\( μ \)) selected by the dose-finding method (MCPMod, MCPMod variants, or MHMC method). To make the summary statistic dimensionless and to account for the “oracle” estimation error we consider

\[
rCAPE = \frac{CAPE}{CAPE_{oracle}}
\]

where \( CAPE_{oracle} = \int_0^1 |μ_0(x, \hat{θ}_0) - μ_0(x, θ_0)| \, dx \), that is the absolute cumulative prediction error obtained when the actual model (\( μ_0 \)) is fitted to the data. For instance, if we simulate a data from the emax model in Table 5, and apply any dose-finding method (say MCPMod) to this data and obtain linlog as the best fitted model for this dose-response data. Then \( μ_0 \) denote the emax model, \( μ \) denote the linlog model , \( CAPE \) denote the cumulative absolute deviation of \( μ \) from \( μ_0 \) and \( CAPE_{oracle} \) denotes the cumulative absolute deviation between the fitted and observed response when we fit the data based on the actual model (\( μ_0 \)).

The above measure \( rCAPE \) is evaluated for two sample sizes: 10 and 25. We do not present results for the larger sample sizes because the methods are performing quite similar for higher sample sizes in terms of the \( rCAPE \).
3. Dose-selection performance: Once the PoC is established and the model has been selected from the candidate set, the final step of dose-finding studies consists of estimating the target dose(s) of interest. Here, the discussion is restricted to estimation of the Minimum Effective Dose (MED). Following Bretz et al. (2005), we defined the MED as:

$$MED = \arg\min_{d \in (d_0, d_k]} \{ \mu_d > \mu_{d_0} + \Delta \}$$

The MED is estimated by:

$$\hat{MED} = \arg\min_{d \in (d_0, d_k]} \{ \hat{\mu}_d > \hat{\mu}_{d_0} + \Delta, L_d > \hat{\mu}_{d_0} \}$$

with $L_d$ the lower $1 - 2\eta$ confidence limit for the mean value $\mu_d$ at dose $d$. A typical choice of $\eta$ is 0.1. $\Delta$ is the clinical relevance threshold that is typical suggested by the physicians. Here, we choose $\Delta = 0.4$ (similar to what used in Bretz et al., 2005).

To evaluate the dose-selection performance of the different dose-finding methods, the relative deviation ($R_i$) of the estimated MED ($\hat{MED}$) from the actual MED is computed in each simulation run $i$:

$$R_i = 100 \frac{\hat{MED}_i - MED}{MED}$$

We present the mean, median and interquartile range of the $R_i$ and $|R_i|$ from 10,000 simulation runs under each simulation scenario. These two measures characterize the bias and precision of $\hat{MED}_i$, respectively. The results are shown for sample size $n_i = 10, 25, 50,$ and $75$ per dose group and $\eta = 0.1$ (corresponding to 80% level confidence intervals).

It should be noted here that the MCP-Mod and the MCP-Mod variants differ only in the PoC testing step but are very similar with respect to their dose-response profile estimation and dose selection step. In the original MCP-Mod method though parameter values need to be specified in the PoC testing step, they are reestimated after the model selection step. So conditional on the same selected model, the methods should give identical results. Thus differences only arise from the fact that sometimes different models are selected as the best one by the different methods. One can have a clear notion of how often the methods select different models from the model selection table (see Table 10).

3.3 | Simulation results

The result of the simulation study are summarized here using the performance metrics of Section 3.2.

1. Detecting Dose Response Trends (DRT): The power to detect the dose-response trend for the candidate sets 1, 2, and 3, respectively are shown in Tables 6, 7, and 8. These three tables only show the $P(DRT)$ for data simulated from models not included in the candidate set of MCP-Mod in each case. For data simulated from nonconstant models included in the candidate set, the results are in favor of the MCP-Mod method as expected. They are shown in tables attached in the Supporting Information. An additional table, Table 9 is also added along with the tables to report which method is performing best in dose-response trend detection in each candidate set and under each data generating shape.

The constant model is included to assess the performance of the four methods in terms of controlling the Type 1 error. It is observed in Tables 6, 7, and 8 that the Type 1 error is properly controlled for the MCP-Mod, MCLRe, and MHMC but it gets inflated for MCLRa for small sample sizes (see Table 6) and for large candidate sets (see Table 7). Thus FWER control might be an issue for the MCLRa method if sample size is small or if the candidate set is large in a dose-response study. For models included in the candidate set of MCP-Mod, power to detect the trend is better (by 2–4%) for MCP-Mod as compared to MCLRa and MCLRe (see Tables 1, 2, and 3 attached in the supporting information). This is obvious because MCP-Mod is using the exact population parameter estimates in their candidate set. Of course, this situation is expected to be rarely the case in practice. What is interesting to observe here, is that the MHMC method has comparable or better power than the MCP-Mod method, even though the population parameters are unknown in this method.

For population shapes not included in the candidate set of MCP-Mod, the performance of MCP-Mod is expected to deteriorate but we observe in Table 9 that it still performs better than the other methods in many situations. For instance, for the linlog data generating shape MCP-Mod can detect the dose response trend with candidate set 3 as good as the other methods even though linlog is not included in its' candidate set (see Table 3). This may be attributed to the fact that the linlog and emax data-generating shape used here (see Figure 1 and Table 5) are close to each other and contrasts corresponding to emax dose response models are able to capture the linlog dose-response trend sufficiently well. Similarly the better performance of MCP-Mod to capture the tlog dose-response trend is attributed to the closeness of the tlog shape with the exponential shape.
TABLE 6  Power to detect the dose response trend under constant model and models not included in the candidate set (Candidate Set 1) of MCP-Mod

| Population shape | Size | MCP-Mod | MCLRa | MCLRe | MHMC (w un) | MHMC (w/o un) |
|------------------|------|---------|-------|-------|-------------|---------------|
| constant         | 10   | 0.05    | 0.06  | 0.05  | 0.05        | 0.05          |
|                  | 25   | 0.05    | 0.05  | 0.05  | 0.04        | 0.05          |
|                  | 50   | 0.05    | 0.05  | 0.05  | 0.04        | 0.05          |
|                  | 75   | 0.05    | 0.05  | 0.05  | 0.05        | 0.05          |
|                  | 100  | 0.05    | 0.05  | 0.05  | 0.05        | 0.05          |
|emax2             | 10   | 0.10    | 0.13  | 0.11  | 0.10        | 0.11          |
|                  | 25   | 0.14    | 0.18  | 0.17  | 0.14        | 0.15          |
|                  | 50   | 0.20    | 0.27  | 0.26  | 0.21        | 0.23          |
|                  | 75   | 0.27    | 0.37  | 0.36  | 0.28        | 0.30          |
|                  | 100  | 0.33    | 0.45  | 0.44  | 0.34        | 0.37          |
|sigmoidal Emax    | 10   | 0.31    | 0.30  | 0.27  | 0.31        | 0.33          |
|                  | 25   | 0.58    | 0.54  | 0.52  | 0.58        | 0.61          |
|                  | 50   | 0.84    | 0.81  | 0.80  | 0.84        | 0.86          |
|                  | 75   | 0.95    | 0.93  | 0.92  | 0.95        | 0.95          |
|                  | 100  | 0.98    | 0.98  | 0.97  | 0.98        | 0.98          |
|power             | 10   | 0.25    | 0.26  | 0.23  | 0.24        | 0.26          |
|                  | 25   | 0.47    | 0.46  | 0.44  | 0.44        | 0.47          |
|                  | 50   | 0.72    | 0.70  | 0.69  | 0.70        | 0.72          |
|                  | 75   | 0.86    | 0.84  | 0.84  | 0.84        | 0.86          |
|                  | 100  | 0.94    | 0.92  | 0.92  | 0.92        | 0.93          |
|tlog              | 10   | 0.26    | 0.25  | 0.23  | 0.24        | 0.26          |
|                  | 25   | 0.48    | 0.46  | 0.45  | 0.44        | 0.47          |
|                  | 50   | 0.76    | 0.73  | 0.72  | 0.71        | 0.73          |
|                  | 75   | 0.89    | 0.87  | 0.87  | 0.85        | 0.87          |
|                  | 100  | 0.95    | 0.95  | 0.94  | 0.93        | 0.94          |
|quadratic         | 10   | 0.16    | 0.18  | 0.16  | 0.15        | 0.16          |
|                  | 25   | 0.30    | 0.31  | 0.29  | 0.25        | 0.27          |
|                  | 50   | 0.50    | 0.51  | 0.49  | 0.42        | 0.43          |
|                  | 75   | 0.66    | 0.66  | 0.65  | 0.56        | 0.57          |
|                  | 100  | 0.77    | 0.77  | 0.77  | 0.68        | 0.68          |

that is included in its' candidate set. But when we deviate drastically from the dose-response shapes included in the candidate set of MCP-Mod, like the second emax shape (emax 2), the power drops by $2 - 11\%$ for MCP-Mod as compared to the other methods. The MCLRa method outperforms the other methods across all the candidate sets for this data generating shape. For the quadratic shape, the power to detect the dose-response trend is better for the MCLRa method compared to the other methods. We observed a power loss by $4 - 11\%$ for the MHMC method with respect to the MCLRa method under this data generating shape. Moreover, we observed that we did not gain much in power from the inclusion of the unrestricted model. Another striking observation that is uniform across all the comparisons and all the data generating shapes is that, the MHMC method detects the dose-response trend better for small sample sizes ($n_i = 10$).

2. Estimating the dose-response profile: The model selection probability tables for candidate set 1 and 2 are attached in the Supporting Information and the model selection probabilities for candidate set 3 is shown in Table 10. Table 10 shows that the MHMC method can identify the linear dose-response shape better than the other methods. But there is no significant difference between the methods for the other data generating shapes. One interesting observation from the model selection probability table is: The Emax dose-response shape is often misclassified as linlog when candidate set 1 and 2 (where both
### Table 7: Power to detect the dose response trend under constant model and models not included in the candidate set (Candidate Set 2) of MCP-Mod

| Population shape | Methods          | Size | MCP-Mod | MCLRa | MHMC (w un) | MHMC (w/o un) |
|------------------|------------------|------|---------|-------|-------------|---------------|
| Constant         |                  | 10   | 0.05    | 0.06  | 0.05        | 0.05          |
|                  |                  | 25   | 0.05    | 0.06  | 0.04        | 0.05          |
|                  |                  | 50   | 0.05    | 0.05  | 0.04        | 0.05          |
|                  |                  | 75   | 0.05    | 0.05  | 0.05        | 0.05          |
|                  |                  | 100  | 0.05    | 0.06  | 0.05        | 0.05          |
| Emax2            |                  | 10   | 0.09    | 0.13  | 0.10        | 0.11          |
|                  |                  | 25   | 0.13    | 0.19  | 0.14        | 0.15          |
|                  |                  | 50   | 0.20    | 0.27  | 0.21        | 0.23          |
|                  |                  | 75   | 0.26    | 0.37  | 0.28        | 0.30          |
|                  |                  | 100  | 0.32    | 0.44  | 0.34        | 0.37          |
| Power            |                  | 10   | 0.25    | 0.27  | 0.24        | 0.26          |
|                  |                  | 25   | 0.47    | 0.47  | 0.44        | 0.47          |
|                  |                  | 50   | 0.72    | 0.71  | 0.70        | 0.72          |
|                  |                  | 75   | 0.86    | 0.85  | 0.84        | 0.86          |
|                  |                  | 100  | 0.94    | 0.93  | 0.92        | 0.93          |
| Tlog             |                  | 10   | 0.25    | 0.26  | 0.24        | 0.26          |
|                  |                  | 25   | 0.48    | 0.46  | 0.44        | 0.47          |
|                  |                  | 50   | 0.75    | 0.72  | 0.71        | 0.73          |
|                  |                  | 75   | 0.88    | 0.87  | 0.85        | 0.87          |
|                  |                  | 100  | 0.95    | 0.94  | 0.93        | 0.94          |
| Quadratic        |                  | 10   | 0.16    | 0.21  | 0.15        | 0.16          |
|                  |                  | 25   | 0.30    | 0.35  | 0.25        | 0.27          |
|                  |                  | 50   | 0.50    | 0.56  | 0.42        | 0.43          |
|                  |                  | 75   | 0.66    | 0.71  | 0.56        | 0.57          |
|                  |                  | 100  | 0.77    | 0.82  | 0.68        | 0.68          |

Emax and linlog are included (see model selection probability tables in Supporting Information). This is mainly because the model selection is based on the AIC criteria, so models with less parameters are more preferred if the model shapes are similar.

To assess the methods in scenarios where the "correct" dose response shape is outside the candidate set, it is necessary to quantify the model selection property of the different methods by another parametric quantity. Thus, the measure $rCAPE$ was introduced in Section 3.2. It is a reasonable measure for assessing the accuracy in estimating the dose-response curve under the different methods. The box plots of $rCAPE$ of the different methods for candidate set 3 and for two sample sizes (10 and 25) are shown in Figure 2. It shows there is no striking difference between the methods for the emax and sigmoidal $Emax$ shape. However, the performance of the MHMHC method is better for dose-response profile estimation under the linear dose-response curve that is also evident from the earlier section. The distribution is comparable for all the methods, with slightly less biased estimates with the MCLRa method. The distribution of $rCAPE$ for tlog is similar to the distribution of $rCAPE$ for sigmoidal $Emax$ justifying the frequent misclassification of the tlog models into sigmoidal $Emax$ shape in Table 10. Thus it is evident here that even if the correct model is not included in the candidate set there may be situations when the approximated fitted model performs as good as the correct model.

MCLRa and MCLRe performed very similar but the computation time of the MCLRa method is much faster (see Table 4). It performs well for small sizes despite being an asymptotic method so it is preferred over MCLRe method. It is hard to choose between MCLRa and MHMC method in terms of characterizing the dose-response profile.

### 3. Dose Selection Performance:

The box plot distribution of MED estimated under the different methods are shown in Figure 3. The relative bias and precision in the MED estimates (measured in terms of $R_i$ and $|R_i|$) described in Section 3.2 are shown...
| Population shape | Size | MC-PMod | MCLRa | MHMC (w un) | MHMC (w/o un) |
|------------------|------|---------|-------|-------------|---------------|
| Constant         | 10   | 0.05    | 0.06  | 0.05        | 0.05          |
|                  | 25   | 0.05    | 0.05  | 0.04        | 0.05          |
|                  | 50   | 0.05    | 0.05  | 0.04        | 0.05          |
|                  | 75   | 0.05    | 0.05  | 0.05        | 0.05          |
|                  | 100  | 0.05    | 0.05  | 0.05        | 0.05          |
| Linlog           | 10   | 0.26    | 0.27  | 0.25        | 0.28          |
|                  | 25   | 0.50    | 0.47  | 0.47        | 0.50          |
|                  | 50   | 0.77    | 0.72  | 0.73        | 0.76          |
|                  | 75   | 0.89    | 0.86  | 0.87        | 0.88          |
|                  | 100  | 0.95    | 0.94  | 0.94        | 0.95          |
| Exponential      | 10   | 0.22    | 0.24  | 0.23        | 0.25          |
|                  | 25   | 0.42    | 0.43  | 0.43        | 0.45          |
|                  | 50   | 0.67    | 0.69  | 0.68        | 0.71          |
|                  | 75   | 0.82    | 0.84  | 0.83        | 0.85          |
|                  | 100  | 0.91    | 0.92  | 0.91        | 0.93          |
| Emax2            | 10   | 0.10    | 0.12  | 0.10        | 0.11          |
|                  | 25   | 0.14    | 0.18  | 0.14        | 0.15          |
|                  | 50   | 0.21    | 0.26  | 0.21        | 0.23          |
|                  | 75   | 0.28    | 0.35  | 0.28        | 0.31          |
|                  | 100  | 0.35    | 0.43  | 0.34        | 0.38          |
| Power            | 10   | 0.25    | 0.26  | 0.24        | 0.27          |
|                  | 25   | 0.48    | 0.46  | 0.44        | 0.48          |
|                  | 50   | 0.74    | 0.70  | 0.70        | 0.73          |
|                  | 75   | 0.87    | 0.84  | 0.84        | 0.86          |
|                  | 100  | 0.94    | 0.92  | 0.92        | 0.94          |
| Tlog             | 10   | 0.22    | 0.24  | 0.23        | 0.25          |
|                  | 25   | 0.42    | 0.44  | 0.43        | 0.46          |
|                  | 50   | 0.67    | 0.71  | 0.69        | 0.71          |
|                  | 75   | 0.82    | 0.86  | 0.84        | 0.85          |
|                  | 100  | 0.90    | 0.94  | 0.92        | 0.93          |
| Quadratic        | 10   | 0.18    | 0.20  | 0.15        | 0.17          |
|                  | 25   | 0.33    | 0.34  | 0.25        | 0.28          |
|                  | 50   | 0.54    | 0.55  | 0.43        | 0.45          |
|                  | 75   | 0.69    | 0.70  | 0.57        | 0.59          |
|                  | 100  | 0.80    | 0.81  | 0.70        | 0.71          |

In Table 11 and Table 6 (in Supporting Information), respectively. Both the table and the plots indicate the dependence of the MED precision on the dose response profile; the exponential, linlog, quadratic, and power shape leads to considerably more biased and less precise MED estimates, for all the methods. Surprisingly, under the tlog shape the MED estimates are accurate even though tlog is not included in the candidate set 3. This might be because the sigEmax shape approximates the tlog shape sufficiently well as shown in Table 10. For dose-response shapes except linear and quadratic shapes, MCLRa method give less biased MED estimates compared to the other methods (see Table 11). MCLRa method estimates the MED with better precision for all the nonlinear dose-response shapes for small sample sizes (see Table 11). However, for large sample sizes the performance of the methods are similar. The better performance of MCLRa method for small sample sizes
### Table 9
Methods selected under different candidate sets and different data generating shapes

| Candidate Set 1 | Candidate Set 2 | Candidate Set 3 |
|-----------------|-----------------|-----------------|
| linlog          | exponential     | power           | sigmoidal |
| emax2           | MCLRa           | MHMC            | MHMC/MCP-Mod/MCLRa |
| power           | MHMC/MCP-Mod    | MHMC/MCP-Mod/MCLRa | MHMC/MCP-Mod |
| tlog            | MCP-Mod         | MHMC            | MHMC/MCP-Mod/MCLRa |
| quadratic       | MCP-Mod/MCLRa   | MCLRa           | MCLRa |

### Table 10
Model selection probabilities under different data generating shapes for the different methods in candidate set 3. Probabilities for selecting the correct model have been given in boldface. Each number is based on 10,000 simulations.

| Population Shape | Selected Model for MCP-Mod | Selected Model for MCLRa | Selected Model for MHMC |
|------------------|----------------------------|--------------------------|-------------------------|
| constant         |                            |                          |                         |
| n                | const | linear | emax | sigE | const | linear | emax | sigE | const | linear | emax | sigE |
| 10               | 0.95  | 0.03   | 0.01 | 0.01 | 0.94  | 0.02   | 0.02 | 0.02 | 0.95  | 0.04   | 0.01 | 0.00 |
| 25               | 0.95  | 0.03   | 0.01 | 0.01 | 0.94  | 0.01   | 0.02 | 0.02 | 0.95  | 0.04   | 0.01 | 0.00 |
| 50               | 0.95  | 0.03   | 0.01 | 0.01 | 0.94  | 0.01   | 0.02 | 0.02 | 0.95  | 0.04   | 0.01 | 0.00 |
| 75               | 0.95  | 0.03   | 0.02 | 0.01 | 0.95  | 0.01   | 0.02 | 0.02 | 0.95  | 0.04   | 0.01 | 0.00 |
| 100              | 0.95  | 0.03   | 0.01 | 0.01 | 0.95  | 0.01   | 0.02 | 0.02 | 0.95  | 0.04   | 0.01 | 0.00 |
| linear           | 0.23  | 0.61   | 0.10 | 0.06 | 0.25  | 0.53   | 0.10 | 0.12 | 0.21  | 0.66   | 0.08 | 0.05 |
| n                | const | linear | emax | sigE | const | linear | emax | sigE | const | linear | emax | sigE |
| 10               | 0.24  | 0.52   | 0.19 | 0.06 | 0.26  | 0.44   | 0.20 | 0.11 | 0.24  | 0.56   | 0.15 | 0.05 |
| 25               | 0.02  | 0.85   | 0.08 | 0.06 | 0.02  | 0.83   | 0.08 | 0.07 | 0.02  | 0.85   | 0.07 | 0.06 |
| 50               | 0.00  | 0.87   | 0.07 | 0.07 | 0.00  | 0.87   | 0.07 | 0.07 | 0.00  | 0.87   | 0.07 | 0.07 |
| 75               | 0.00  | 0.86   | 0.07 | 0.07 | 0.00  | 0.86   | 0.07 | 0.07 | 0.00  | 0.86   | 0.07 | 0.07 |
| 100              | 0.00  | 0.86   | 0.07 | 0.07 | 0.00  | 0.86   | 0.07 | 0.07 | 0.00  | 0.86   | 0.07 | 0.07 |
| linlog           | 0.43  | 0.43   | 0.43 | 0.43 | 0.04  | 0.04   | 0.04 | 0.04 | 0.04  | 0.04   | 0.04 | 0.04 |
| n                | const | linear | emax | sigE | const | linear | emax | sigE | const | linear | emax | sigE |
| 10               | 0.26  | 0.32   | 0.32 | 0.32 | 0.27  | 0.41   | 0.27 | 0.06 | 0.27  | 0.41   | 0.27 | 0.06 |
| 25               | 0.02  | 0.38   | 0.51 | 0.51 | 0.02  | 0.36   | 0.51 | 0.11 | 0.02  | 0.38   | 0.49 | 0.11 |
| 50               | 0.00  | 0.19   | 0.69 | 0.69 | 0.00  | 0.19   | 0.69 | 0.12 | 0.00  | 0.19   | 0.68 | 0.13 |
| 75               | 0.00  | 0.09   | 0.77 | 0.77 | 0.00  | 0.09   | 0.77 | 0.13 | 0.00  | 0.09   | 0.77 | 0.14 |
| 100              | 0.00  | 0.05   | 0.82 | 0.82 | 0.00  | 0.05   | 0.82 | 0.14 | 0.00  | 0.05   | 0.81 | 0.14 |
| exponential      | 0.79  | 0.08   | 0.13 | 0.13 | 0.71  | 0.05   | 0.20 | 0.05 | 0.76  | 0.10   | 0.14 | 0.01 |
| n                | const | linear | emax | sigE | const | linear | emax | sigE | const | linear | emax | sigE |
| 10               | 0.79  | 0.08   | 0.13 | 0.13 | 0.71  | 0.05   | 0.20 | 0.05 | 0.76  | 0.10   | 0.14 | 0.01 |
| 25               | 0.59  | 0.10   | 0.30 | 0.30 | 0.01  | 0.47   | 0.41 | 0.04 | 0.54  | 0.12   | 0.33 | 0.01 |
| 50               | 0.34  | 0.09   | 0.56 | 0.56 | 0.01  | 0.20   | 0.71 | 0.03 | 0.26  | 0.10   | 0.63 | 0.01 |
| 75               | 0.18  | 0.06   | 0.75 | 0.75 | 0.01  | 0.08   | 0.86 | 0.02 | 0.11  | 0.06   | 0.81 | 0.01 |
| 100              | 0.09  | 0.04   | 0.86 | 0.86 | 0.01  | 0.02   | 0.03 | 0.03 | 0.04  | 0.04   | 0.91 | 0.01 |
| emax2            | 0.12  | 0.65   | 0.10 | 0.13 | 0.15  | 0.58   | 0.10 | 0.17 | 0.13  | 0.69   | 0.08 | 0.10 |
| n                | const | linear | emax | sigE | const | linear | emax | sigE | const | linear | emax | sigE |
| 10               | 0.12  | 0.66   | 0.08 | 0.25 | 0.00  | 0.66   | 0.08 | 0.26 | 0.00  | 0.66   | 0.08 | 0.26 |
| 25               | 0.00  | 0.46   | 0.07 | 0.47 | 0.00  | 0.46   | 0.07 | 0.47 | 0.00  | 0.46   | 0.07 | 0.47 |
| 50               | 0.00  | 0.32   | 0.06 | 0.62 | 0.00  | 0.32   | 0.06 | 0.62 | 0.00  | 0.32   | 0.06 | 0.62 |
| 75               | 0.00  | 0.21   | 0.05 | 0.74 | 0.00  | 0.21   | 0.05 | 0.74 | 0.00  | 0.21   | 0.05 | 0.74 |
| 100              | 0.00  | 0.21   | 0.05 | 0.74 | 0.00  | 0.21   | 0.05 | 0.74 | 0.00  | 0.21   | 0.05 | 0.74 |
may be because of the type 1 error inflation. An interesting observation noted is that for nonlinear shapes like power and sigEmax, the MED estimated by fitting the “correct” model is more biased compared (see Table 11) to the MED estimated under the different dose-finding methods. This further supports our argument that predicting the “correct” dose-response shape might not be always necessary. Our goal is achieved if we get a dose response shape that can well approximate the right model.
FIGURE 3  The box plot distribution of estimated MED under different methods and across all different data generating shapes in candidate set 3. The horizontal line indicates the real MED from the simulated model and the “Actual” box plot distributions shows the MED estimates obtained when the “correct” model is fitted on the simulated data.

4 | DISCUSSION

This section summarizes the conclusions from the simulation studies. We have observed from our simulations that though MCP-Mod have more constraints compared to the other methods yet it is performing well in many situations. However, the performance of the MCP-Mod is highly influenced by its choice of candidate set. While the MCP-Mod variant, MHMC method performs uniformly better across all the simulations and under different choice of candidate set for monotone dose-response
Table 11: Median, mean, and interquartile range (IQR) of the relative deviation ($R_i$) from the target dose estimated under the different methods and different dose-response shapes with candidate set 3. Results are given for the “Actual” case that is when the correct model is fitted on the simulated data and when the model selected under different dose-finding methods are fitted to the simulated data.

| Population shape | MED Size Actual | MCP-Mod Actual | MCLRa Actual | MHMC Actual | Mean | MCP-Mod | MCLRa | MHMC | IQR |
|------------------|----------------|----------------|--------------|-------------|------|----------|-------|------|-----|
| Linear           | 10 -0.01 -0.18 | -0.21 -0.14    | 0.04 -0.21   | -0.25 -0.16 | 0.59 | 0.38     | 0.36  | 0.42 |
|                  | 25 -0.00 -0.04 | -0.04 -0.03    | 0.04 -0.03   | -0.04 -0.01 | 0.35 | 0.36     | 0.35  | 0.37 |
|                  | 50 0.00 -0.02  | -0.02 -0.02    | 0.03 -0.02   | -0.02 -0.01 | 0.25 | 0.27     | 0.27  | 0.27 |
|                  | 75 0.00 -0.02  | -0.02 -0.02    | 0.02 -0.02   | -0.02 -0.02 | 0.20 | 0.22     | 0.22  | 0.22 |
| emax             | 10 -0.10 0.49  | -0.14 0.70     | 0.25 0.39    | 0.21 0.53   | 1.51 | 1.97     | 1.83  | 2.17 |
|                  | 25 -0.00 0.01  | -0.04 -0.02    | 0.28 0.53    | 0.50 0.61   | 1.29 | 2.09     | 2.06  | 2.10 |
|                  | 50 -0.01 -0.09 | -0.09 -0.09    | 0.21 0.28    | 0.28 0.51   | 0.98 | 1.15     | 1.14  | 1.14 |
|                  | 75 0.00 -0.08  | -0.08 -0.08    | 0.17 0.18    | 0.18 0.21   | 0.82 | 0.76     | 0.76  | 0.75 |
| Sigmoidal Emax   | 10 -0.16 0.06  | 0.01 0.09      | -0.11 0.04   | -0.02 0.09  | 0.71 | 0.52     | 0.48  | 0.56 |
|                  | 25 -0.08 0.12  | 0.12 0.12      | -0.06 0.12   | 0.12 0.13   | 0.55 | 0.52     | 0.52  | 0.52 |
|                  | 50 -0.04 -0.09 | 0.09 0.09      | -0.04 0.05   | 0.05 0.05   | 0.45 | 0.48     | 0.48  | 0.48 |
|                  | 75 -0.01 0.07  | 0.07 0.07      | -0.03 0.02   | 0.02 0.02   | 0.38 | 0.42     | 0.42  | 0.42 |
| linlog           | 10 0.00 0.09   | 0.01 0.15      | 0.20 -0.01   | -0.10 0.06  | 1.16 | 0.90     | 0.93  | 0.76 |
|                  | 25 -0.00 0.26  | 0.26 0.26      | 0.15 0.17    | 0.15 0.20   | 0.65 | 1.01     | 1.00  | 1.01 |
|                  | 50 -0.00 0.23  | 0.23 0.23      | 0.09 0.13    | 0.13 0.14   | 0.44 | 0.88     | 0.88  | 0.88 |
|                  | 75 0.00 0.10   | 0.10 0.10      | 0.07 0.08    | 0.08 0.11   | 0.36 | 0.79     | 0.79  | 0.79 |
| Power            | 10 -0.16 0.13  | 0.03 0.20      | -0.07 0.00   | -0.10 0.08  | 1.30 | 1.13     | 1.11  | 1.19 |
|                  | 25 -0.06 0.32  | 0.31 0.33      | 0.00 0.20    | 0.18 0.25   | 0.97 | 1.24     | 1.23  | 1.25 |
|                  | 50 -0.02 0.25  | 0.24 0.24      | 0.03 0.15    | 0.15 0.19   | 0.76 | 1.08     | 1.08  | 1.08 |
|                  | 75 0.00 0.04   | 0.04 0.03      | 0.04 0.11    | 0.11 0.12   | 0.62 | 0.95     | 0.95  | 0.95 |
| Exponential      | 10 0.02 -0.30  | -0.30 -0.27    | -0.04 -0.30  | -0.28 -0.26 | 0.23 | 0.28     | 0.37  | 0.33 |
|                  | 25 0.02 -0.13  | -0.13 -0.13    | -0.02 -0.13  | -0.13 -0.11 | 0.16 | 0.32     | 0.32  | 0.33 |
|                  | 50 0.01 0.11   | -0.11 0.11     | -0.01 -0.11  | -0.11 -0.10 | 0.13 | 0.25     | 0.25  | 0.25 |
|                  | 75 0.00 0.10   | -0.10 0.10     | -0.01 -0.10  | -0.10 0.11  | 0.21 | 0.21     | 0.21  | 0.21 |
| Quadratic        | 10 -0.11 0.35  | -0.53 -0.49    | 0.12 0.06    | -0.21 0.16  | 0.61 | 1.30     | 0.75  | 2.01 |
|                  | 25 -0.05 0.17  | -0.20 -0.20    | 0.06 0.28    | 0.18 0.47   | 0.46 | 1.07     | 0.91  | 0.92 |
|                  | 50 -0.02 -0.14 | -0.14 -0.14    | 0.06 0.15    | 0.13 0.69   | 0.37 | 0.71     | 0.70  | 0.71 |
|                  | 75 -0.01 0.11  | -0.11 0.11     | 0.05 0.10    | 0.10 1.24   | 0.31 | 0.57     | 0.57  | 0.57 |
| tlog             | 10 -0.14 0.23  | -0.21 -0.20    | -0.13 -0.23  | -0.20 0.19  | 0.37 | 0.31     | 0.48  | 0.39 |
|                  | 25 -0.02 0.01  | -0.01 -0.01    | -0.06 0.04   | -0.04 0.02  | 0.34 | 0.33     | 0.33  | 0.33 |
|                  | 50 -0.01 0.01  | 0.01 0.01      | 0.05 -0.01   | -0.01 0.01  | 0.34 | 0.27     | 0.27  | 0.27 |
|                  | 75 0.00 0.02   | 0.02 0.02      | -0.05 0.00   | 0.00 0.00   | 0.34 | 0.23     | 0.23  | 0.23 |

In terms of computation time, the MCLR method is better than the other methods.
It is worth mentioning here that for the simulation studies conducted above, despite the small differences, the methods are highly comparable and for most of the scenarios the model selection probability or the dose-selection power differed only by 3–4%. So it is hard to select or argue in favor of any particular method. It is also important to note that MCP-Mod, MCLRa, and MCLRe differs essentially in the PoC testing step while the model selection and dose selection procedure are similar for all of them. Other model selection criteria like BIC and likelihood ratios are also explored but there is no significant improvement observed over the AIC criterion. Also note that the MHMC method is the only method where candidate models are compared with each other in a multiple testing framework, not just with the constant model.

The benefits of MCP-Mod is that it is already extended to general parametric setting by Pinheiro, Bornkamp, Glimm, and Bretz (2014). The optimal design consideration and determination of appropriate sample size for the MCP-Mod approach is also suggested in Pinheiro et al. (2006a). Using similar procedure suggested in Pinheiro et al. (2006a) and Dette, Bretz, Pepelyshev, and Pinheiro (2008) sample size can be estimated for the new approaches. Power and sample size computation might be a bit challenging for the MCLRa and MHMC approach because of the nonstandard asymptotic results used here. However, Gutjahr and Bornkamp (2017) have shown the computation of rejection probabilities under alternative as well as null distribution for the MCLRe approach. So power computation and sample size estimation at the design stage of the experiment is feasible for this approach. This is of crucial importance for clinical trials.

The MCLRa approach can also be extended to general parametric families and it can be further modified to address heteroscedasticity in the data that is not so obvious in the MCP-Mod approach. The MHMC approach can also be extended to generalized linear model but the results used in this approach needs to be adjusted for non-normal distributed errors. For MCLRe, to incorporate the non-normality in the residuals one needs to extend the results for elliptically contoured distribution which can be obtained following Fang and Zhang (1990) and Gupta, Varga, and Bodnar (2013).

A simple and obvious alternative to the above dose-finding approaches is to conduct linear trend test using contrasts (Tukey, Ciminera, and Heyse, 1985) or to use dose-response modelling under simple a order restriction (Otava et al., 2014). However the LR methodology for detecting linear as well as non-linear trends are more powerful compared to the above approaches and lead to more efficacious modelling under diverse scenarios. Another possible alternative to parametric approach in dose-response studies is using non-parametric approaches. Bornkamp and Ickstadt (2009) proposed a monotone non-parametric regression in a Bayesian framework where they defined the dose-response function by an appropriately scaled mixture of distribution functions. A disadvantage of such a non-parametric approach is that they cannot be applied to clinical data with a small number of dose-groups. Gallant (1977) have also suggested an alternative exact test for testing lack of fit when regularity conditions are violated similar to the PoC testing step of the approaches discussed in this article. But it have been pointed out that this approach is reasonably applicable for non linear models with one parameter but becomes quite awkward in non linear models with more than one parameter. Further, the approaches discussed in this paper could be of interest outside dose-response analysis. Test of trend and subsequently characterizing the correct shape is a major challenge which arises in various domain of applied sciences like economics or genomic studies.

ACKNOWLEDGMENTS
This work is supported by funding from the European Union’s Horizon 2020 research and innovation program IDEAS (“Improving Design, Evaluation and Analysis of early drug development Studies”; www.ideas-itn.eu) under the Marie Sklodowska-Curie grant agreement No 633567. We are also grateful to the reviewers for their valuable comments and feedback.

CONFLICTS OF INTEREST
The authors have declared no conflicts of interest.

ORCID
Saswati Saha http://orcid.org/0000-0002-9157-8275
Werner Brannath http://orcid.org/0000-0002-8622-3904

REFERENCES
Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, 1383–1414.
Baayen, C., Hougaard, P., & Pipper, C. (2015). Testing effect of a drug using multiple nested models for the dose–response. *Biometrics, 71*(2), 417–427.
Bornkamp, B., & Ickstadt, K. (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose–response analysis. *Biometrics, 65*(1), 198–205.

Bretz, F., Pinheiro, J. C., & Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics, 61*(3), 738–748.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics, 57*–578.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika, 74*(1), 33–43.

Dette, H., Bretz, F., Pepelyshev, A., & Pinheiro, J. (2008). Optimal designs for dose-finding studies. *Journal of the American Statistical Association, 103*(483), 1225–1237.

Dette, H., Titoff, S., Volgushev, S., & Bretz, F. (2015). Dose response signal detection under model uncertainty. *Biometrics, 71*(4), 996–1008.

EMA (2014). Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty, Committee for Medicinal Products for Human Use (CHMP). http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2014/02/WC500161027.pdf.

 Fang, K.-T., & Zhang, Y.-T. (1990). *Generalized multivariate analysis.* Berlin, DE: Science Press Beijing and Springer-Verlag.

Committee for Medicinal Products for Human Use (2014). Qualification opinion of mcp-mod as an efficient statistical methodology for model-based design and analysis of phase ii dose finding studies under model uncertainty. Technical report, EMA/CHMP/SAWP/757052/2013.

Gallant, A. R. (1977). Testing a nonlinear regression specification: A nonregular case. *Journal of the American Statistical Association, 72*(359), 523–530.

Gupta, A. K., Varga, T., & Bodnar, T. (2013). *Elliptically contoured models in statistics and portfolio theory.* Berlin, DE: Springer.

Gutjahr, G. & Bornkamp, B. (2017). Likelihood ratio tests for a dose-response effect using multiple nonlinear regression models. *Biometrics, 73*(1), 197–205.

Hothorn, L., Neuhäuser, M., & Koch, H.-F. (1997). Analysis of randomized dose-finding-studies: Closure test modifications based on multiple contrast tests. *Biometrical Journal, 39*(4), 467–479.

Lindsay, B. G. (1995). Mixture models: Theory, geometry and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics, 5*, i–163.

Liu, X., & Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics, 31*(3), 807–832.

Morgan, B. J. (1992). *Analysis of quantal response data* (Vol. 46). Boca Raton, FL: CRC Press.

Otava, M., Shkedy, Z., Lin, D., Goehlmann, H. W., Bijnens, L., Talloen, W., & Kasim, A. (2014). Dose–response modeling under simple order restrictions using bayesian variable selection methods. *Statistics in Biopharmaceutical Research, 6*(3), 252–262.

Pinheiro, J., Bornkamp, B., & Bretz, F. (2006a). Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. *Journal of Biopharmaceutical Statistics, 16*(5), 639–656.

Pinheiro, J., Bornkamp, B., & Bretz, F. (2014). Model-based dose finding under model uncertainty using general parametric models. *Statistics in medicine, 33*(10), 1646–1661.

Pinheiro, J. C., Bretz, F., & Branson, M. (2006b). Analysis of dose–response studies modeling approaches. 146–171.

Ruberg, S. J. (1995). Dose response studies i. some design considerations. *Journal of Biopharmaceutical Statistics, 5*(1), 1–14.

Stewart, W. H., & Ruberg, S. J. (2000). Detecting dose response with contrasts. *Statistics in Medicine, 19*(7), 913–921.

Tamhane, A. C., Hochberg, Y., & Dunnett, C. W. (1996). Multiple test procedures for dose finding. *Biometrics, 21–37.*

Tamhane, A. C., & Logan, B. R. (2002). Multiple test procedures for identifying the minimum effective and maximum safe doses of a drug. *Journal of the American Statistical Association, 97*(457), 293–301.

Tukey, J., Ciminera, J., & Heyse, J. (1985). Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics, 41*, 295–301.

**SUPPORTING INFORMATION**

Additional Supporting Information including source code to reproduce the results may be found online in the supporting information tab for this article.

**How to cite this article:** Saha S, Brannath W. Comparison of different approaches for dose response analysis. *Biometrical Journal.* 2019;61:83–100. [https://doi.org/10.1002/bimj.201700276](https://doi.org/10.1002/bimj.201700276)