Confirmatory testing for a beneficial treatment effect in dose-response studies using MCP-Mod and an adaptive interim analysis

Franz König[1]

[1]Medical University of Vienna, Austria
Center for Medical Statistics, Informatics and Intelligent Systems
Franz.Koenig@meduniwien.ac.at
www.meduniwien.ac.at/user/franz.koenig

Isaac Newton Institute
July 9, 2015

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013–602552.

Acknowledgments of collaborators

- Frank Bretz (Novartis)
- Bjoern Bornkamp (Novartis)
- Ekkehard Glimm (Novartis)
- Alexandra Graf (Medical University of Vienna)
- Sergii Krasnozhon (Medical University of Vienna)
- Gernot Wassmer (Aptiv Solutions, an ICON plc company)

This work has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013–602552 (IDEAL - http://www.ideal.rwth-aachen.de/).

Introduction

- Traditionally two main analysis strategies for phase II trials: multiple comparisons (MCP) of contrasts between doses and modeling of dose response (DR)
- Bretz et al. (2005) proposed an unified approach combining the advantages of MCP and modeling - MCPMod:
  - Set of candidate models to account for model uncertainty
  - Test PoC using MCP with optimal model contrasts
  - Select a model and estimate target doses (e.g., MED, ED50, ...)
  - Qualification Opinion adopted by CHMP (2014)

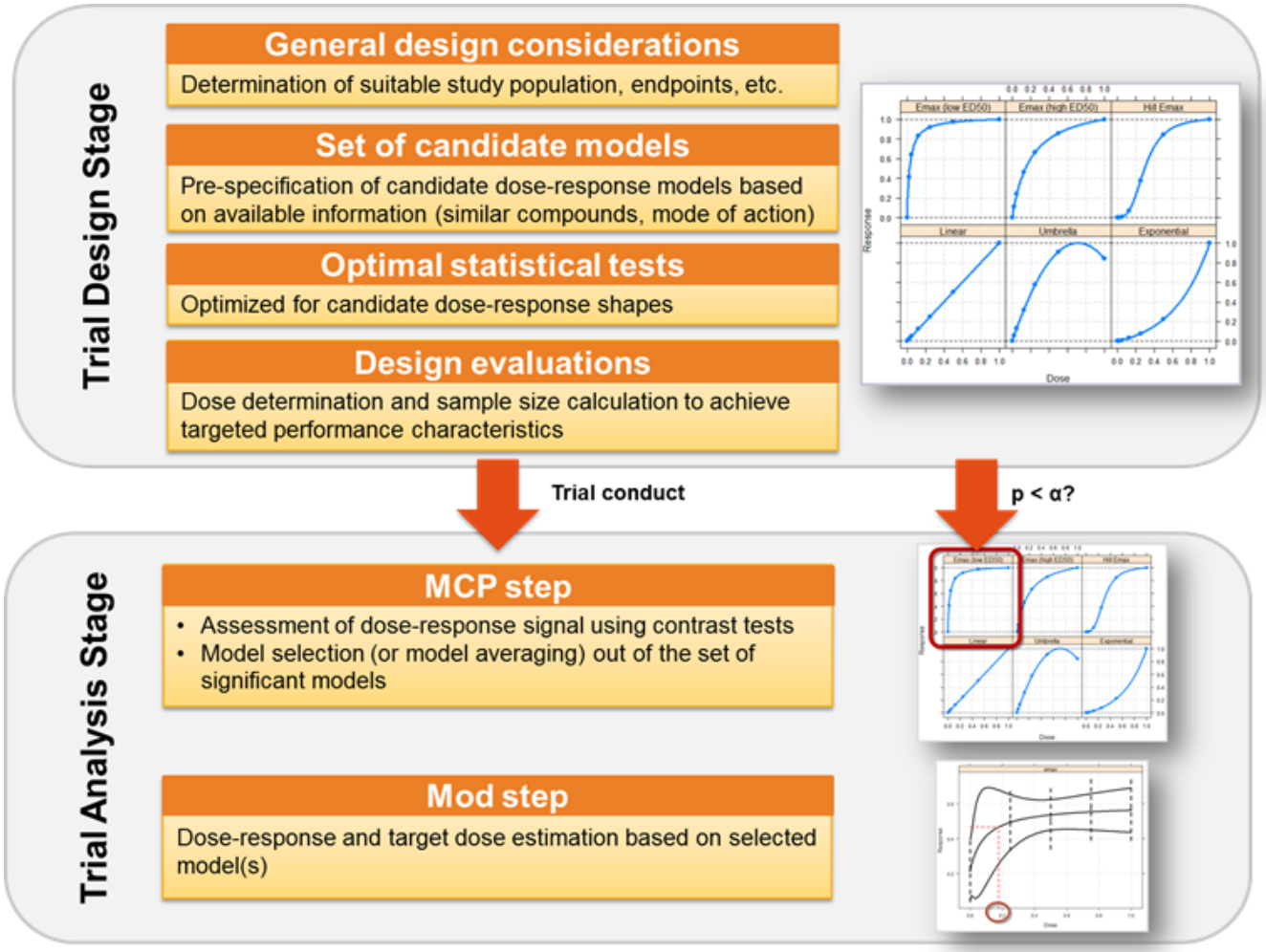
Our goal today:

Use framework of MCPMod to

- allow testing of individual dose-control comparisons
- allow design modifications at an adaptive interim analysis
- increase the power of declaring effective doses statistically significant

Overview of MCP-Mod

(Bretz et al. 2005, Bornkamp et al. 2009, Pinheiro et al. 2014, ...)



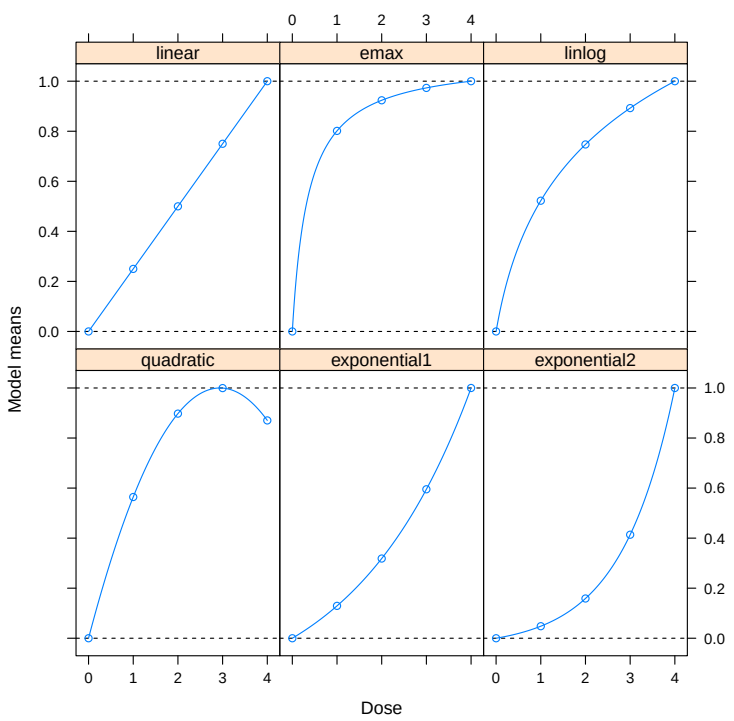
Statistical model

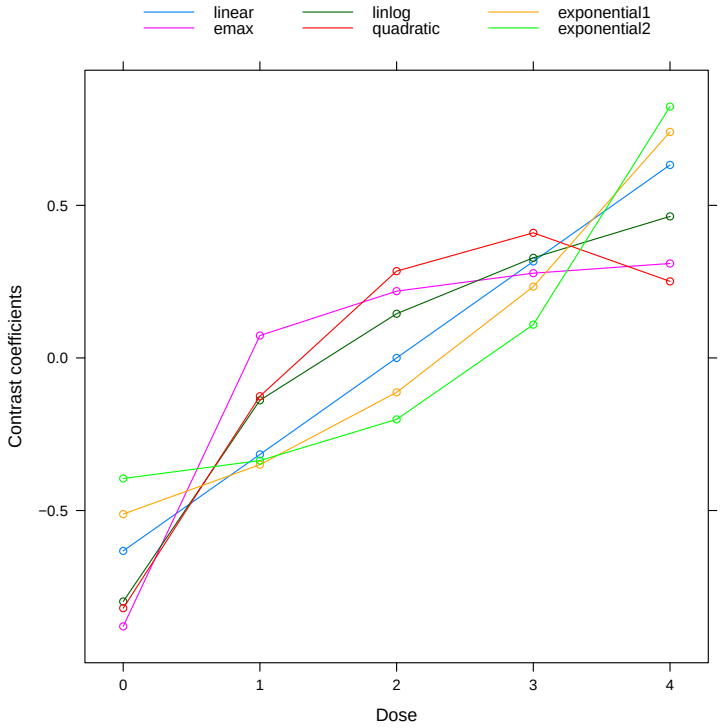
- Assume that a response Y is observed for $k+1$ parallel groups corresponding to doses d_0,d_1,..., d_k (d_0 typically placebo)
- Assume that the Y_{ij} for subject j in group i is modeled as

$$Y_{ij} = f(d_i, \theta) + \epsilon_{ij}, \quad \epsilon_{ij} \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2), \; i = 0, \ldots, k, j = 1, \ldots, n_i,$$

- Denote the true (unknown) effect at dose i with $\mu_i = f(d_i, \theta)$
- MCP-Mod uses classical contrast tests, with multiplicity adjustment, to detect evidence in favor of a dose response signal
  - based on a pre-specified set $\mathcal{M}$ of M parameterized candidate models with corresponding model functions $f_m(d, \theta_m)$, $m = 1, \ldots, M$

Candidate model set





- Use prior guesses to determine the parameters of the standardized models θ_m^0 determining the model shapes
- Optimal contrasts $\mathbf{c}_m = (c_{m0}, c_{m1}, \ldots, c_{mk})'$ depends on best guesses for standardized model.

Contrast test for dose response signal detection (PoC)

- Single contrast test for detecting the mth model shape:

$$T_m = \frac{\sum_{i=0}^{k} c_{mi} \bar{Y}_i}{S\sqrt{\sum_{i=0}^{k} c_{mi}^2 / n_i}}, \quad m = 1, \ldots, M,$$

- Test statistics are jointly multivariate-t distributed with correlations determined by the model contrasts
- Multiplicity adjusted critical value q or adjusted p-values p_m for individual tests of models derived from multivariate t-distribution accounting for the different models
- Dose response signal established if the maximum test statistic $\max_m T_m > q$ or the minimum adjusted p-value $\min_m p_m < \alpha$

Efficacy claims for individual doses using MCP-Mod

- Original MCP-Mod approach is designed for Phase II studies and enables only the detection of a dose response signal at the MCP step
- Often, efficacy claims for individual doses are of interest in Phase III
- Formally interested in testing the k elementary null hypotheses

$$H_i : \mu_i \leq \mu_0, \text{ versus } A_i : \mu_i > \mu_0, \ i = 1, ..., k$$

but within the framework of MCP-Mod

Combine two different methodological concepts

- MCP-Mod contrast test and closed testing principle

For k hypotheses $H_1, \ldots, H_k$:

- Specify a level α test for all $H_J = \bigcap_{i \in J} H_i$, $J \subseteq \{1, \ldots, k\}$, and
- Reject H_i if and only if all H_J with $i \in J$ are rejected

Proposed approach: Details

- Take the original contrast test from MCP-Mod as the level α test for the global intersection hypothesis

$$H_1 \cap H_2 ... \cap H_k : \mu_1, ..., \mu_k \leq \mu_0$$

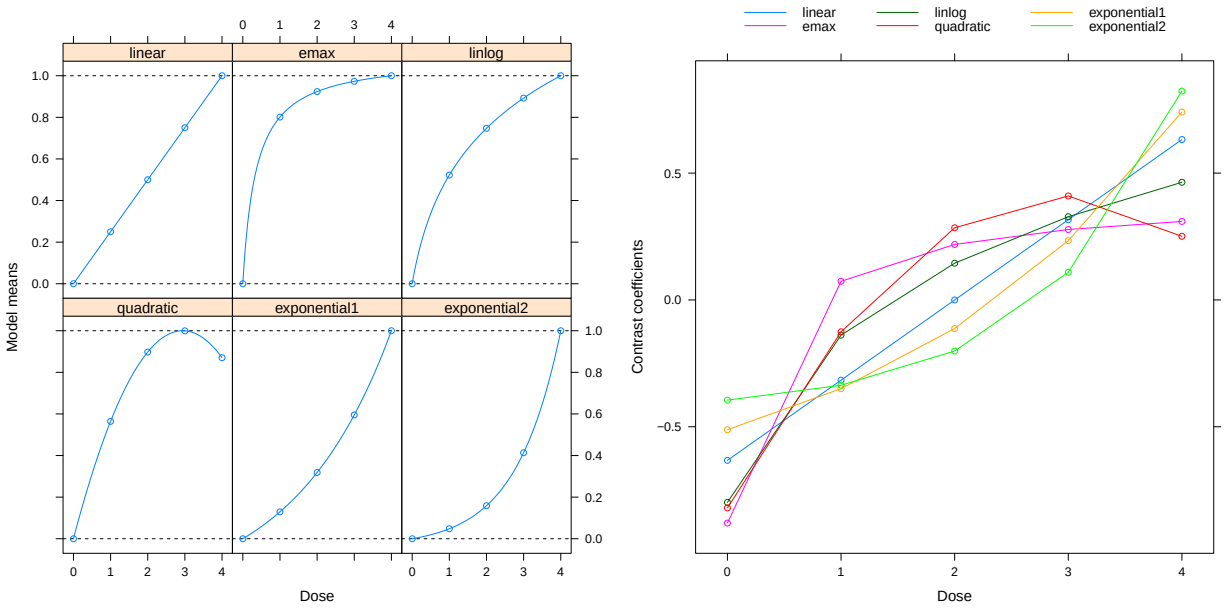
- Use the same strategy for all other intersections hypotheses

Our proposal for a level α test for $H_J = \cap_{i \in J} H_i$, $J \subseteq \{1, \ldots, k\}$

- Test H_J with MCP-Mod contrast tests using doses $i \in J$ only.
- Use the same set of pre-specified set $\mathcal{M}$ of candidates models for each intersection hypothesis H_J, but
- But calculate new optimal contrasts for each H_J and model using only doses $i \in J$
- This means for each intersection hypothesis we have to perform as many contrast tests as models are included in $\mathcal{M}$.

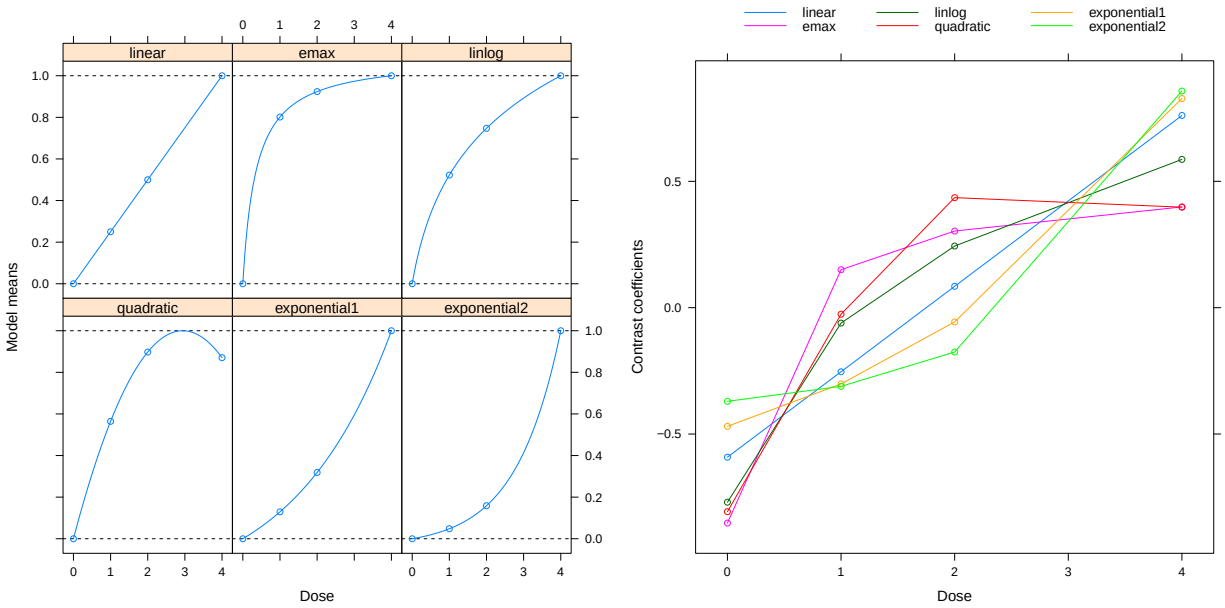
Proposed approach: Illustration of details

- Consider an example with 4 doses and placebo
- Assume 6 candidate models, with corresponding optimal coefficients for testing $H_1 \cap H_2 \cap H_3 \cap H_4 : \mu_1, \mu_2, \mu_3, \mu_4 \leq \mu_0$



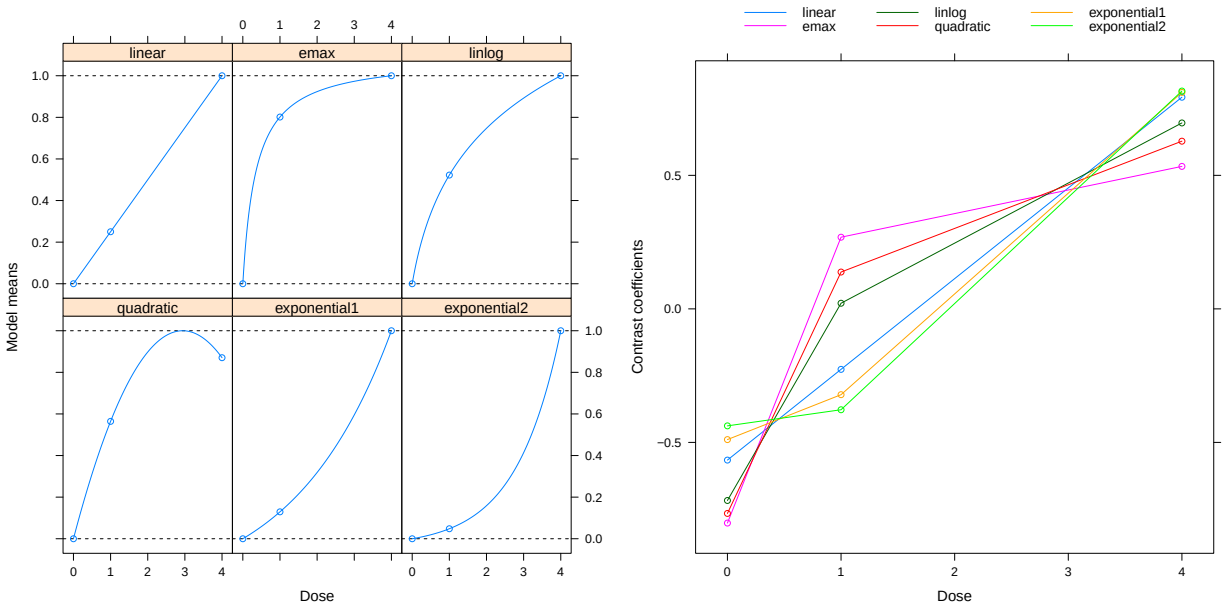
Proposed approach: Illustration of details

- Consider an example with 4 doses and placebo
- Assume 6 candidate models, with corresponding optimal coefficients for testing $H_1 \cap H_2 \cap H_4 : \mu_1, \mu_2, \mu_4, \leq \mu_0$



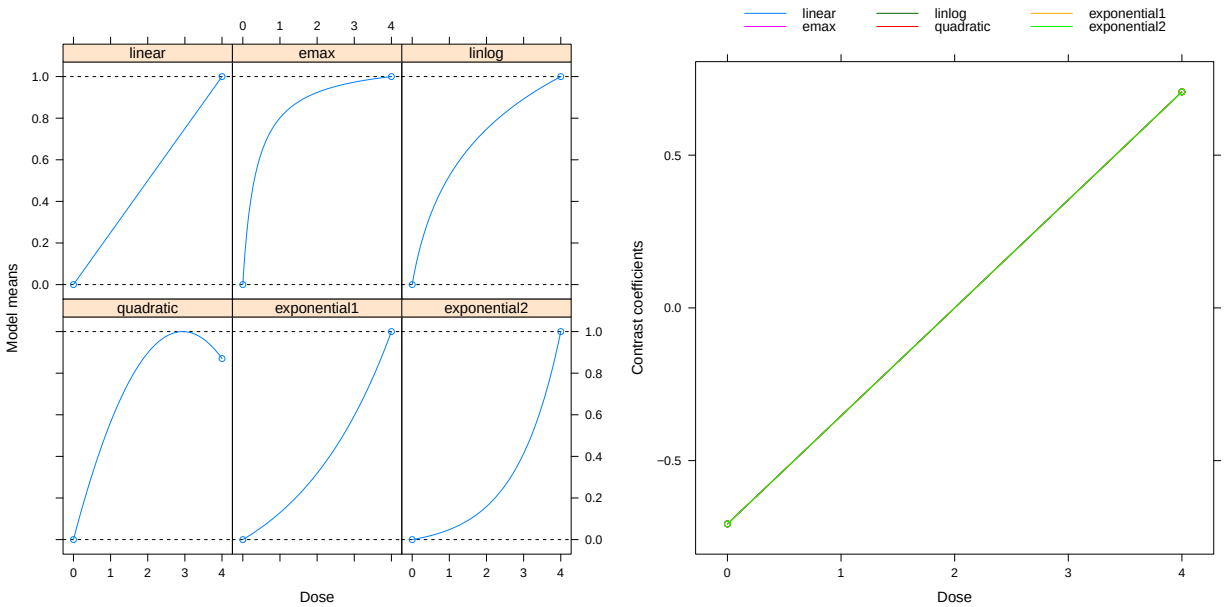
Proposed approach: Illustration of details

- Consider an example with 4 doses and placebo
- Assume 6 candidate models, with corresponding optimal coefficients for testing $H_1 \cap H_4 : \mu_1, \mu_4 \leq \mu_0$



Proposed approach: Illustration of details

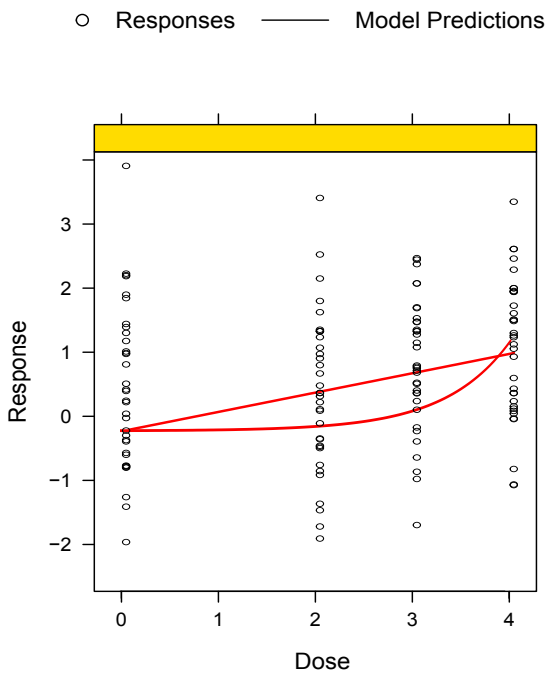
- Consider an example with 4 doses and placebo
- Assume 6 candidate models, with corresponding optimal coefficients for festing $H_4 : \mu_4 \leq \mu_0$



For the test of an elementary null hypothesis this results in the same contrast for all models. This gives a *t*-test at full level α.

Example: Case-study Closed MCP-Mod at $\alpha = 0.025$

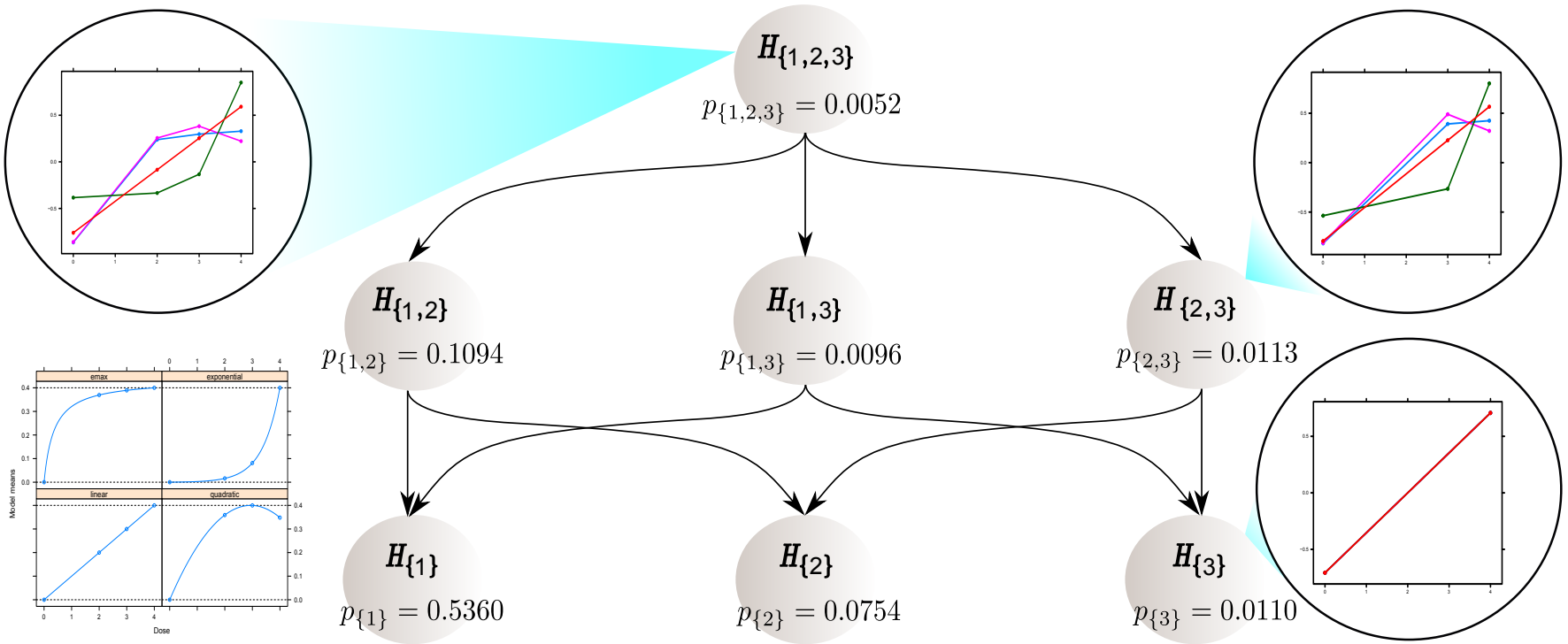
- **control** and $k = 3$ **dose groups** ($d = (0, 2, 3, 4)$);
- **sample size** $N = 140$ (35 patients per arm);
- $\mathcal{M} = \{\texttt{emax}, \texttt{exponential}, \texttt{linear}, \texttt{quadratic}\}$.



Contrast	t-value (T_m)	adj. P-value
exponential	2.790	0.006
linear	2.541	0.011
emax	1.627	0.089
quadratic	1.501	0.111

- PoC can be established ($H_{\{1,2,3\}}$ is **rejected**).

Example: Case-study Closed MCP-Mod



- H_3 **can also be rejected.**

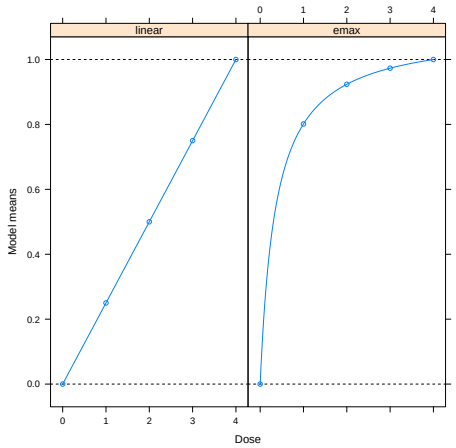
Proposed approach: Remarks

- What if the pre-specified contrasts for some models become very similar for a particular intersection hypothesis H_J?
  - This may e.g. be relevant when the number of doses for some J is small
- Would it not be better to reduce the number of candidate models a-priori for this particular hypothesis to decrease the penalty for multiplicity?
- High similarity between candidate models will result in high correlations, resulting in little multiplicity adjustment, since correlations are accounted for when calculating critical values, p-values, ...

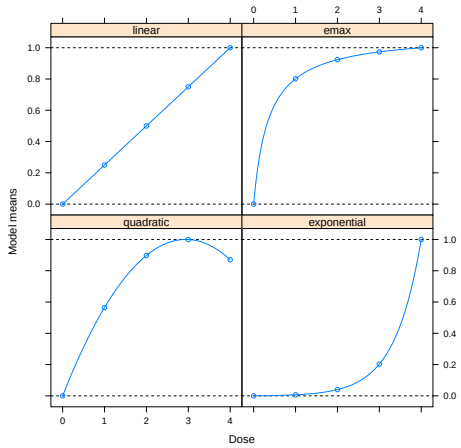
- Question: For each (intersection) hypothesis m contrast tests have to be performed. Is it worth all the effort?

Simulation study

- Parallel group design with $k = 4$ dose groups and placebo, $d_0, d_1, \ldots, d_k$
- Normal responses with variance $\sigma^2 = 1$.
- Testing one sided hypotheses at overall level $\alpha = 0.025$
- Sample size per group $n = 20$
- Two sets of candidate models $\mathcal{M}$:

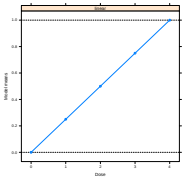


and

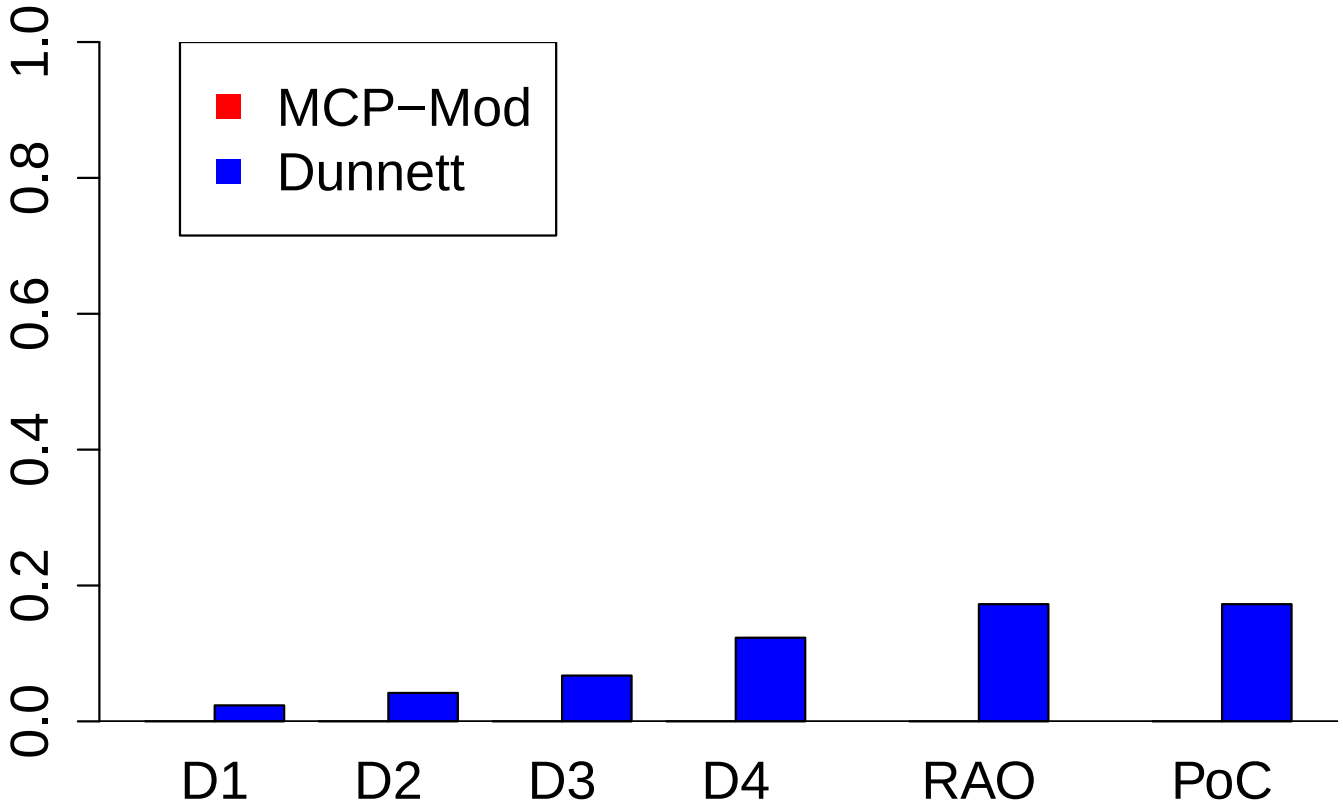


- Compare power closed MCP-Mod with step-down Dunnett test
- 10 000 simulation runs for each parameter configuration

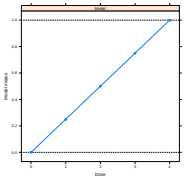
Simulations



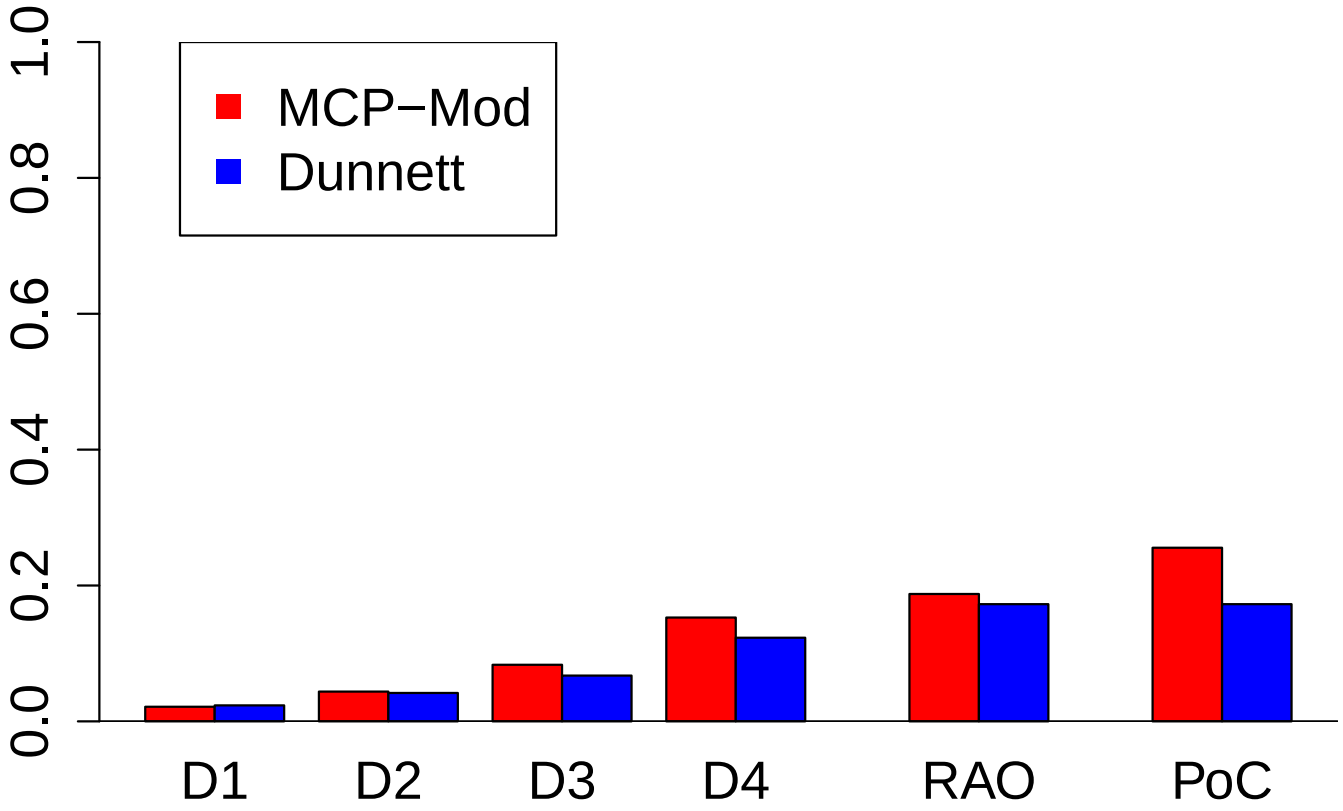
Linear with $\mu_4 = 0.4\sigma$; 2 candidate models



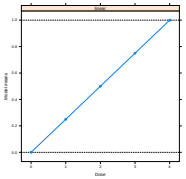
Simulations



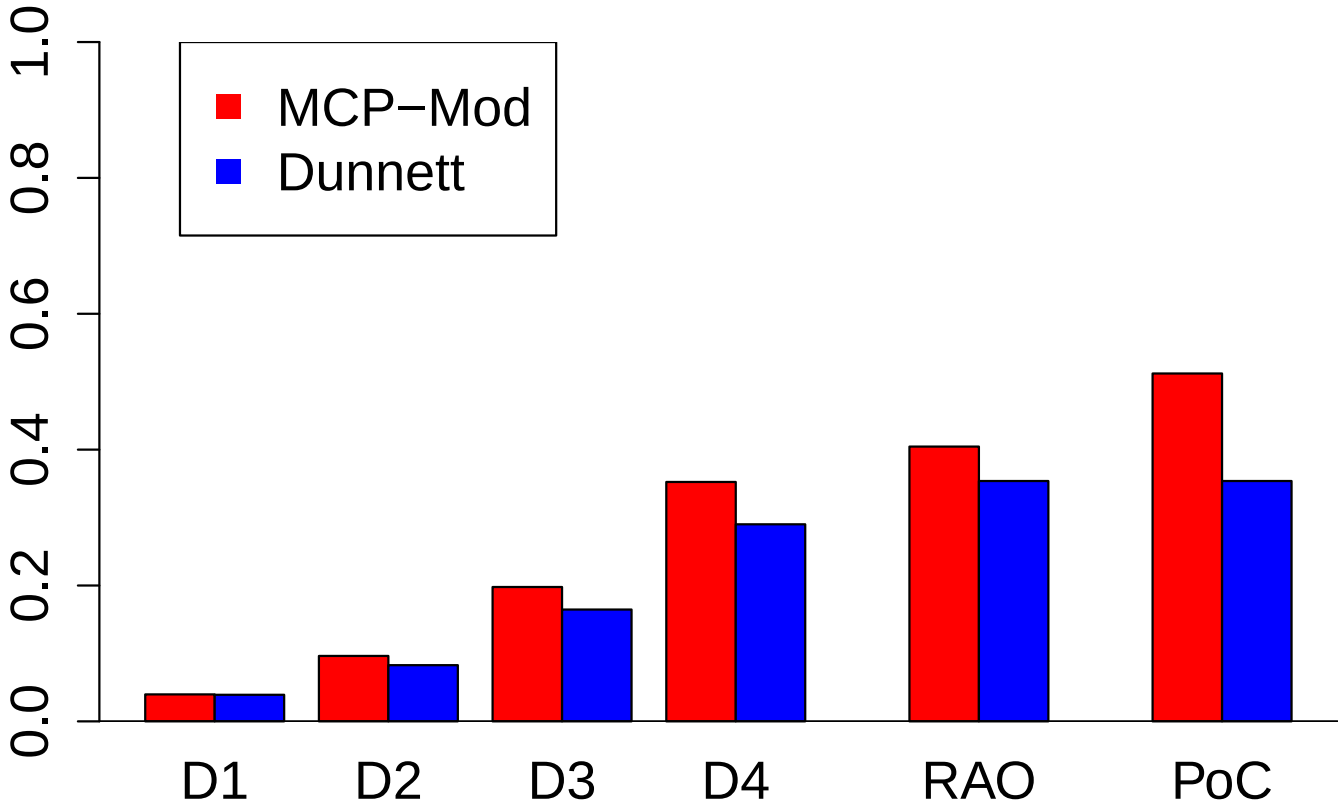
Linear with $\mu_4 = 0.4\sigma$; 2 candidate models



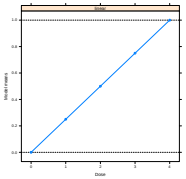
Simulations



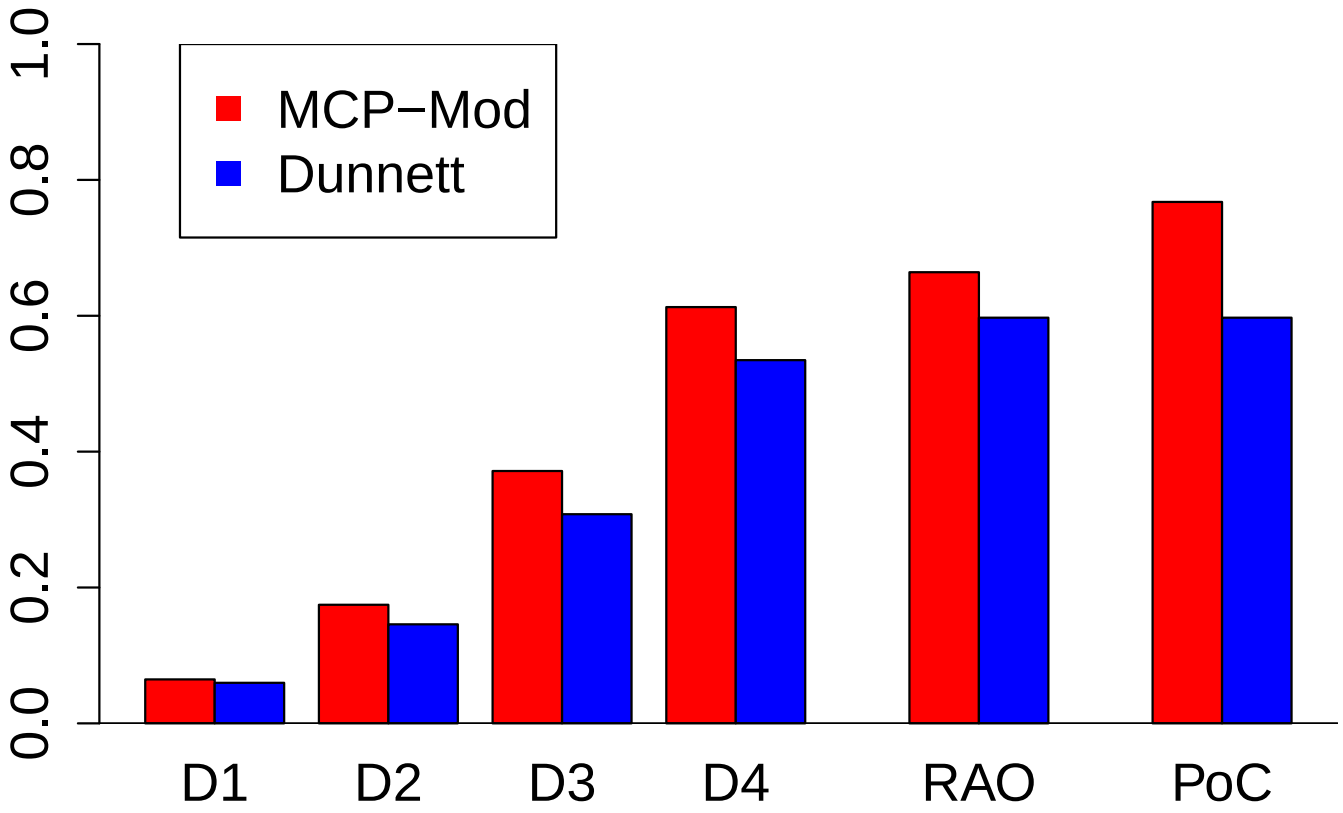
Linear with $\mu_4 = 0.6\sigma$; 2 candidate models



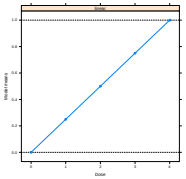
Simulations



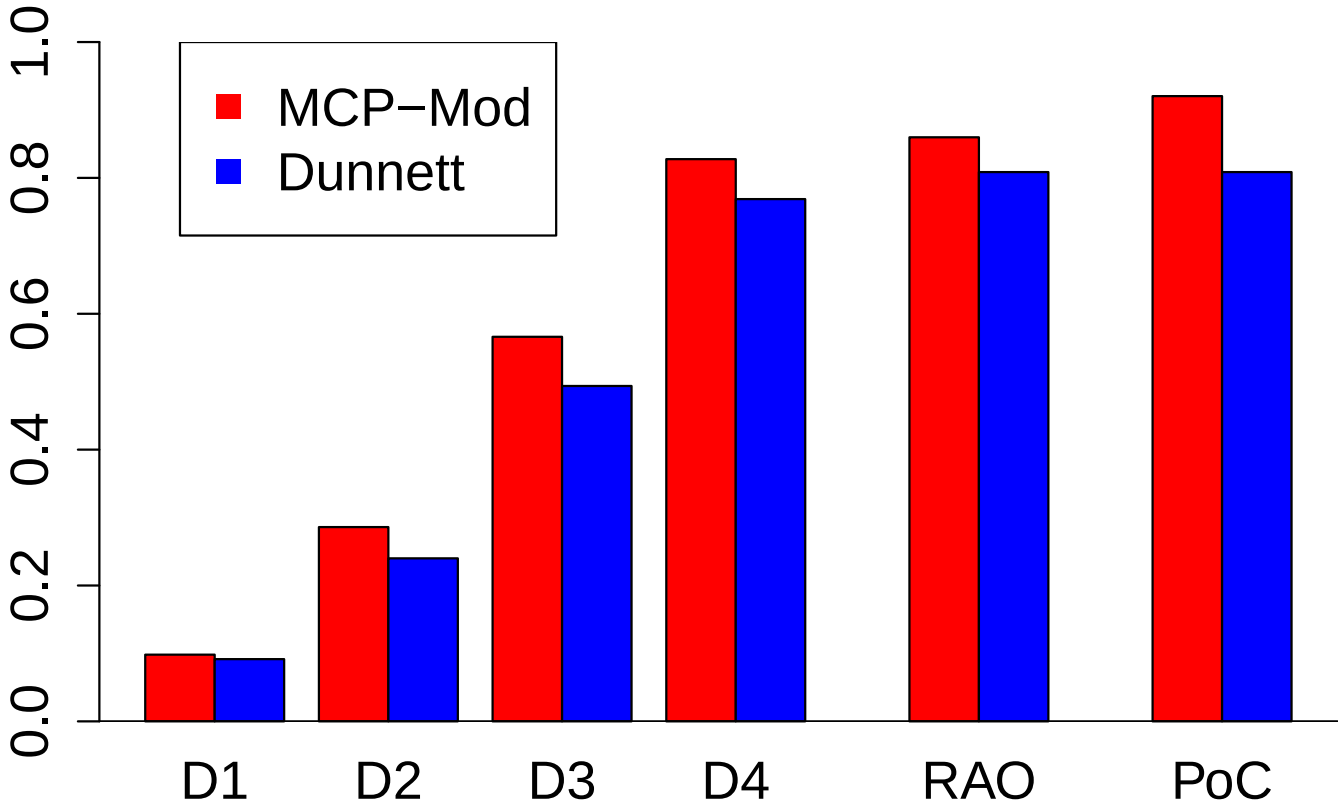
Linear with $\mu_4 = 0.8\sigma$; 2 candidate models



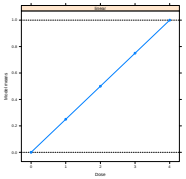
Simulations



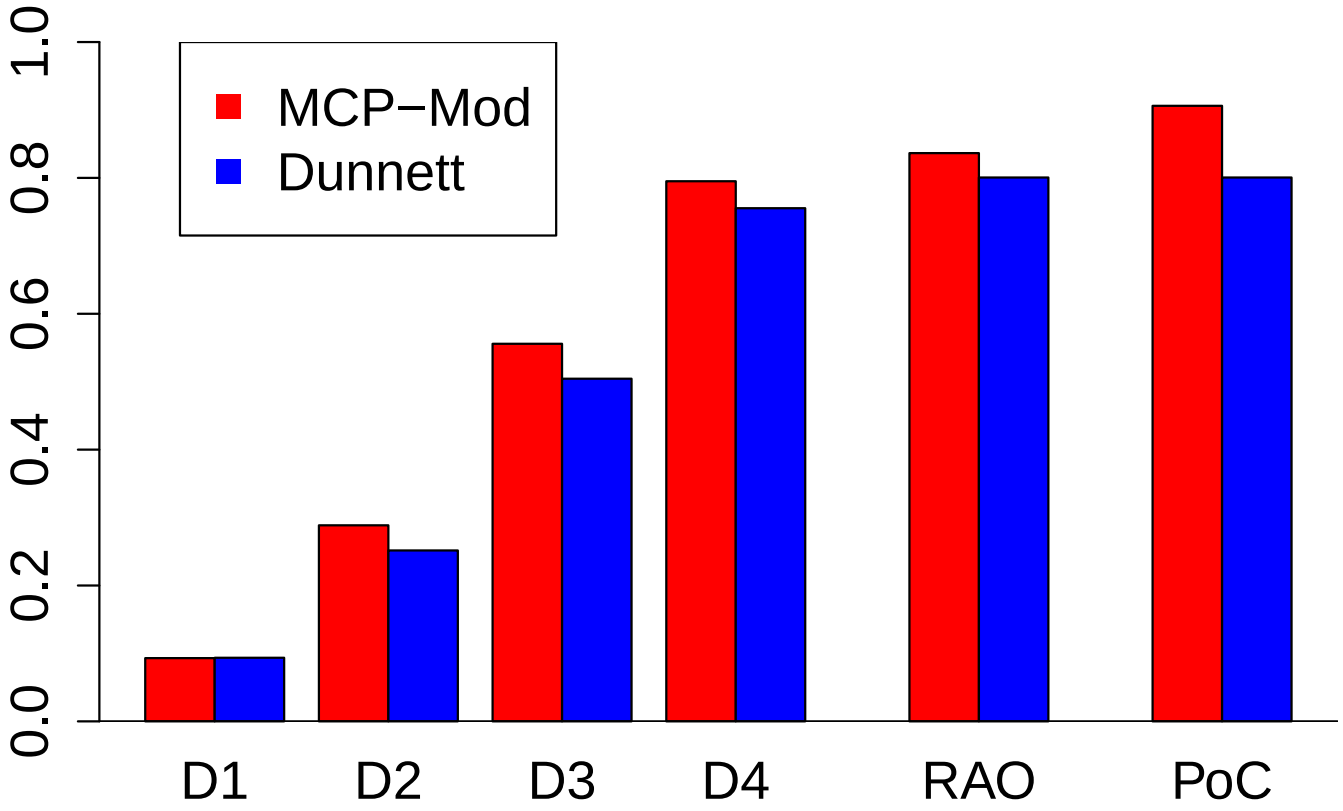
Linear with $\mu_4 = \sigma$; 2 candidate models



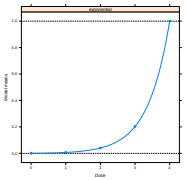
Simulations



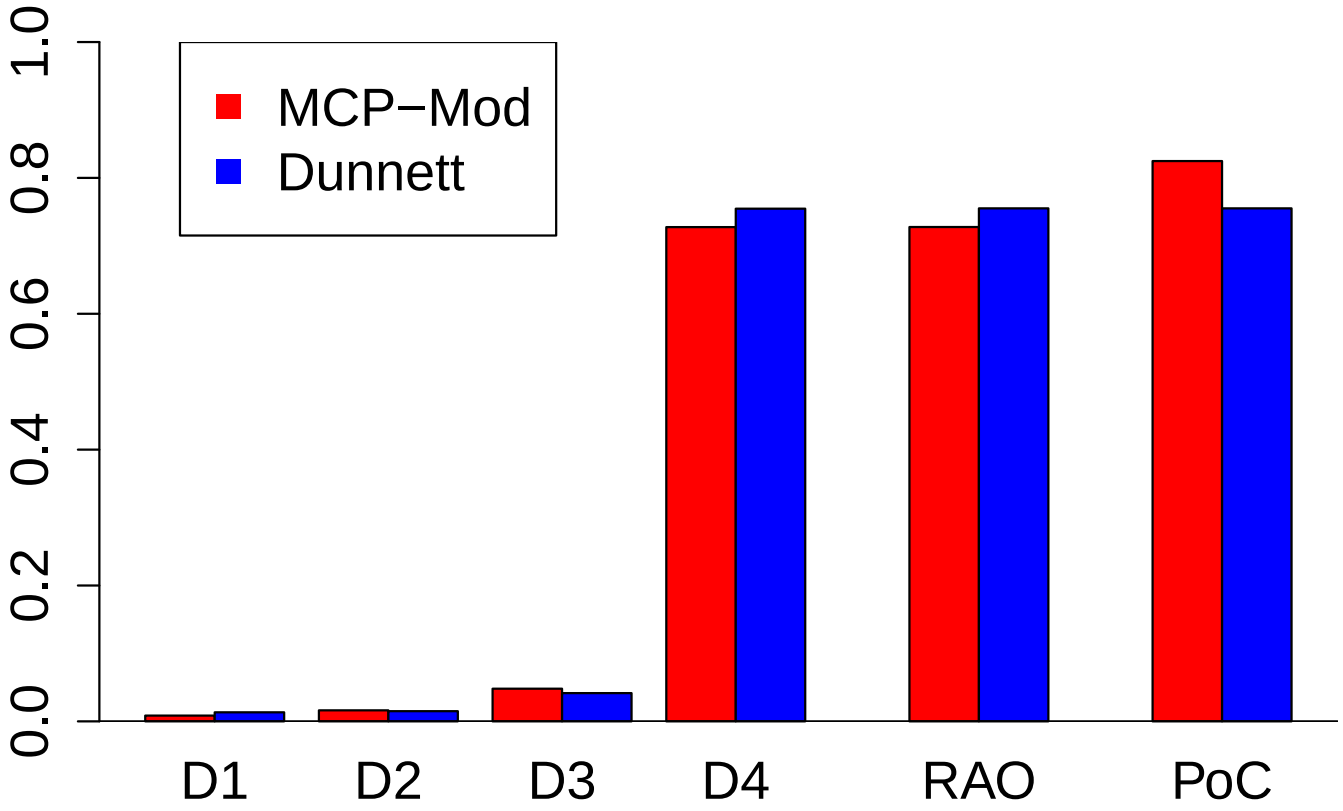
Impact of using more candidates? 4 candidate models



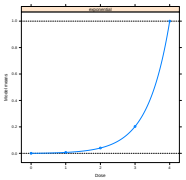
Simulations



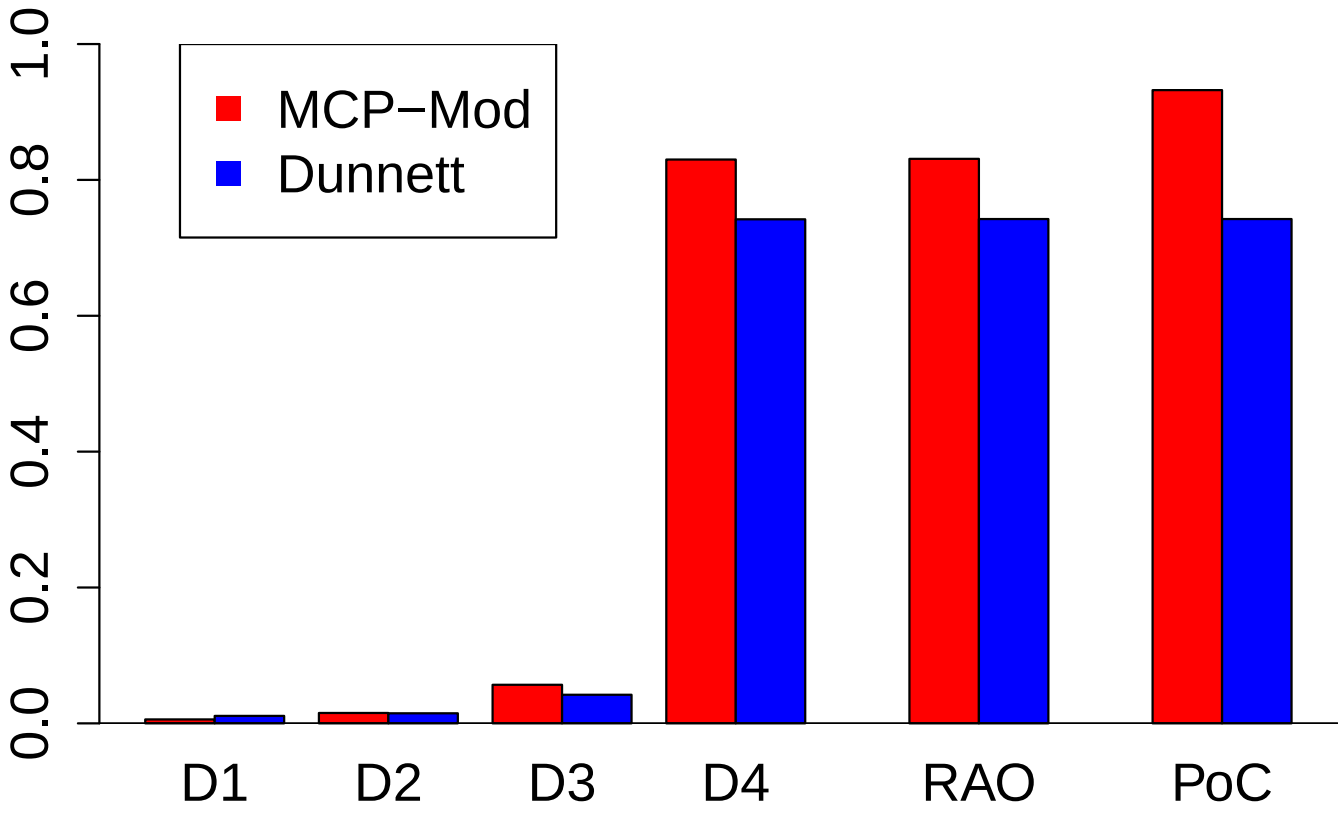
True Model exponential with $\mu_4 = \sigma$; 2 candidate models



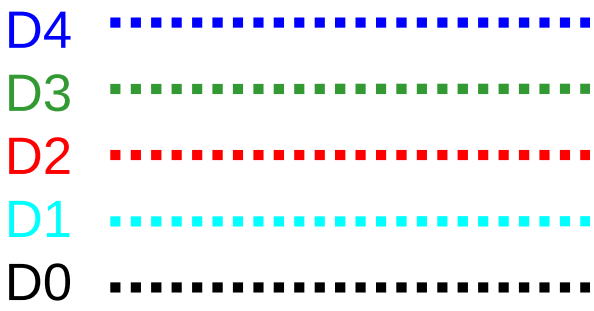
Simulations



True Model exponential with $\mu_4 = \sigma$; 4 candidate models

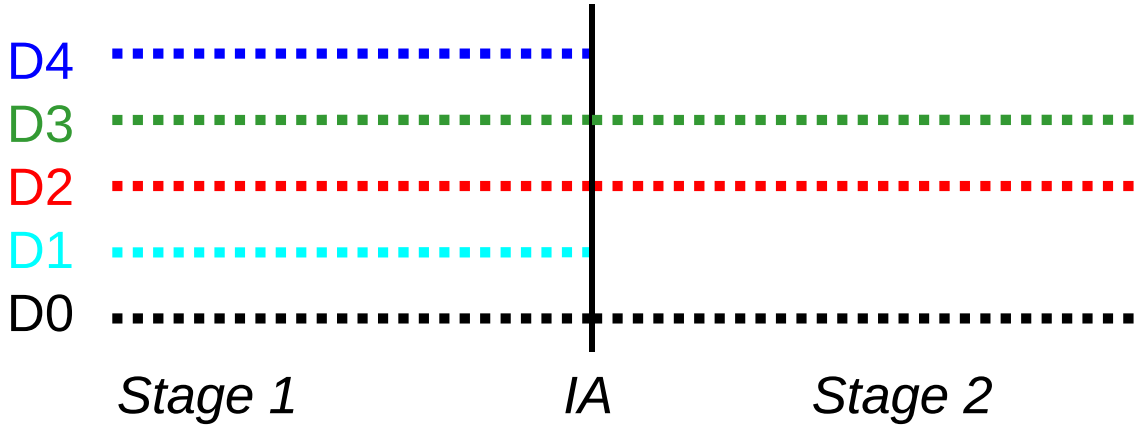


- Fixed Sample Design

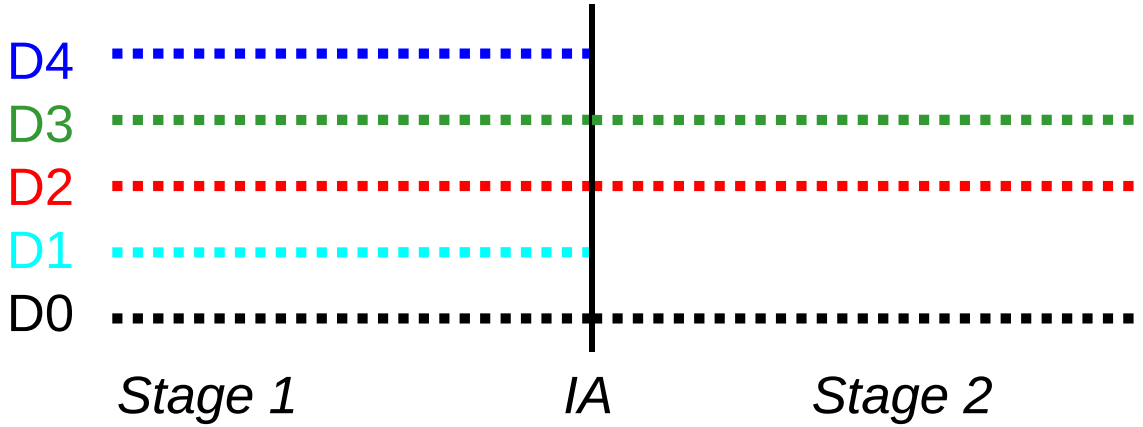


Analysis
(Testing & Estimation)

Adaptive Design



Adaptive Design



- Before trial start:
  - Fix the design of the first stage (sample sizes, dose groups, ...) and a candidate model set $\mathcal{M}_1$ for the first stage.
- At interim:
  - Fix the design of the second stage (adapt sample sizes, selection of doses, ...) and a candidate model set $\mathcal{M}_2$ (drop/add models, refine parameter guesses, ...).
  - Use the Mod part of MCP-Mod to support interim decisions (e.g., estimate MED or the peak dose)
- Final analysis:
  - Test for a dose response signal and/or individual $H_i : \mu_i \leq \mu_0$ using data from both stages with stage-wise tests based on $\mathcal{M}_1$ and $\mathcal{M}_2$

Testing efficacy of selected doses

(BAUER AND KIESER 1999, HOMMEL 2001, POSCH ET AL. 2005, BRETZ ET AL. 2009, ...)

Final test procedure combines three different methodologies:

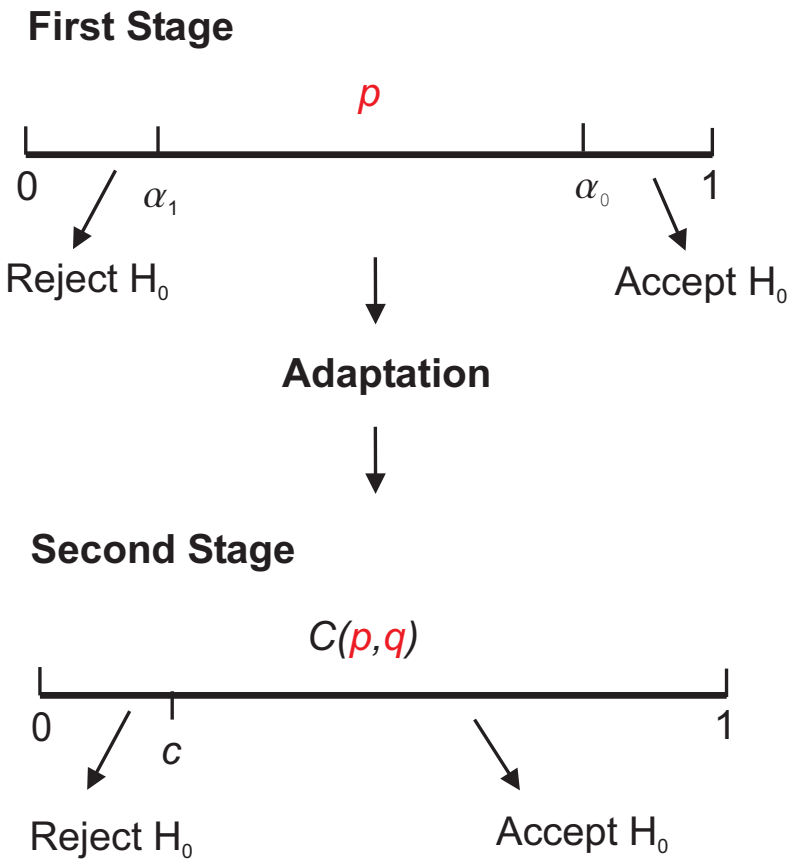
- MCP part of MCP-Mod (above)
- Closed test procedure (above)
- Adaptive combination tests (below) that do not require pre-specified adaptation rules

More specifically, use a combination test $C(p, q)$ to combine the stagewise p-values p and q for each intersection hypothesis and apply the closed test procedure, where

- C could be e.g. Fisher´s product test or the inverse normal method
- p and q are calculated from the extended MCP-Mod approach above

Adaptive Two Stage Test based on Combination Tests

(BAUER 1989, BAUER & KÖHNE 1994, . . .)



Planning:

- Fix design (only) for Stage 1
- Fix combination function $C(p, q)$ and critical value c
  e.g. Inverse-Normal with $C(p, q) = 1 - \Phi[w_1 \Phi^{-1}(1-p) + w_2 \Phi^{-1}(1-q)]$

Stage 1:

- Compute p-value p from Stage 1 data
- Fix design for Stage 2 based on data from Stage 1

Stage 2:

- Compute p-value q form Stage 2 data.
- Reject H_0 iff $C(p, q) \leq c$.

Remarks

- For each intersection hypothesis H_J, $J \subseteq \{1, ..., k\}$ calculate multiplicity adjusted p-values per stage using the set of models $\mathcal{M}_1$ and $\mathcal{M}_2$ for stage 1 and 2, respectively.
- Combine the stage-wise minimum p-values via combination test.
- For stage 2: If for H_J, $J \subseteq \{1, ..., k\}$ doses $i \in J$ are missing at the second stage, simply take the second stage p-value of the largest subset $I \subset J$, where all doses $i \in I$ are available at stage 2.

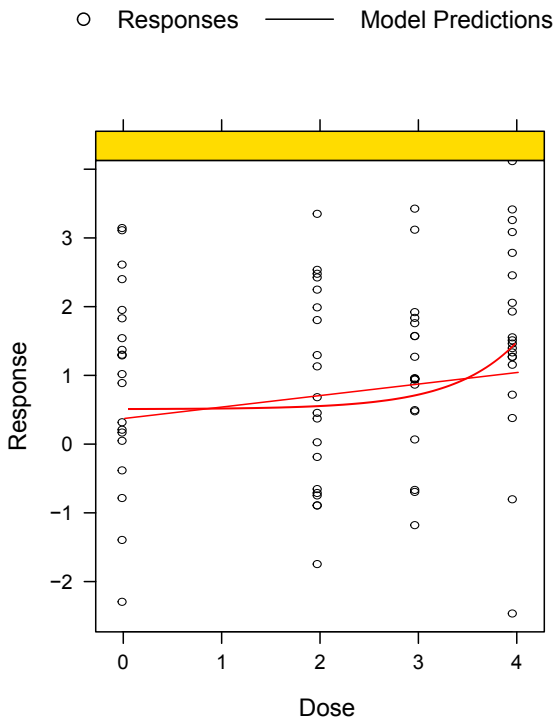
Example: Case-study Adaptive MCP-Mod

Design Considerations

- **control** and $k = 3$ **dose groups** ($d = (0, 2, 3, 4)$);
- multiple level $\alpha = 0.025$ (one-sided);
- 1 adaptive interim analysis with O'Brien Fleming type boundaries ($\alpha_1 = 0.0043$, $\alpha_0 = 0.5$, $c = 0.0235$)
- Inverse Normal
  $C(p, q) = 1 - \Phi[w_1 \Phi^{-1}(1 - p) + w_2 \Phi^{-1}(1 - q)]$, where $w_1^2 = n_1/(n_1 + n_2)$, $w_2^2 = n_2/(n_1 + n_2)$ and n_i is proportional to pre-planned stagewise i sample sizes;
- Stage 1 **sample size** $n_1 = 80$ (20 patients per arm);
- Stage 2 **sample size** $n_2 = 60$
- $\mathcal{M}_1 = \{\texttt{emax}, \texttt{exponential}, \texttt{linear}, \texttt{quadratic}\}$.

Example: Adaptive MCP-Mod (3 doses and control)

- Stage I **sample size** $n_1 = 80$ (20 patients per arm)
- **Interim analysis (Stage I data)**



O'Brien Fleming stopping boundaries

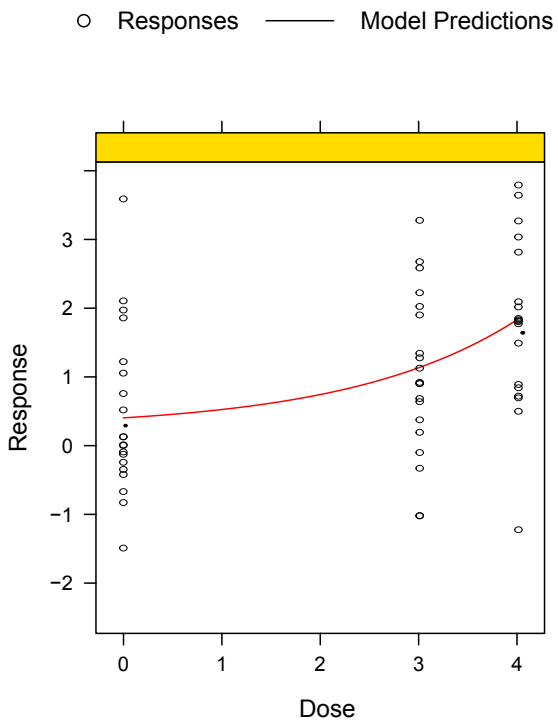
$\alpha_1 = 0.0043$, $\alpha_0 = 0.5$.

Contrast	t-value (T_m)	adj. P-value
exponential	1.758	0.070
linear	1.420	0.129
emax	0.643	0.370
quadratic	0.479	0.435

- Interim Decisions
  - **control** and $k = 2$ **dose groups** ($d = (0, 3, 4)$);
  - Stage II **sample size reallocated** $n_2 = 60$ (20 patients per arm);
  - $\mathcal{M}_2 = \{\texttt{exponential}, \texttt{linear}\}$.

Example: Adaptive MCP-Mod

- **Stage II data**

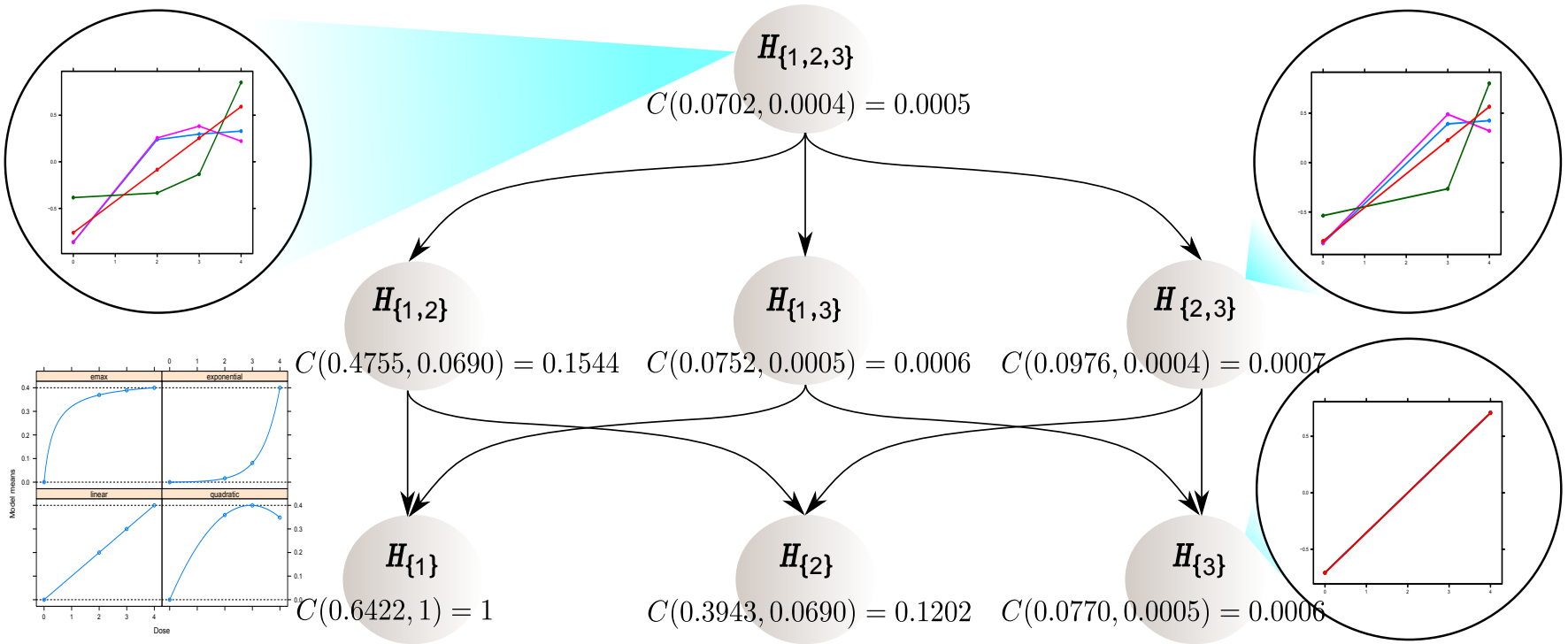


Contrast	t-value (T_m)	adj. P-value
exponential	3.635	0.0004
linear	3.412	0.001

- **Final analysis for PoC**
  - $p = 0.070$;
  - $q = 0.0004$;
  - $C(0.070, 0.0004) = 0.0005 < 0.0235 \Rightarrow H_{\{1,2,3\}}$ is **rejected**.

Example: Adaptive MCP-Mod based on Closed Test

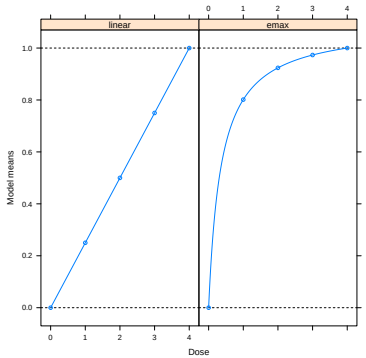
- **Final analysis**



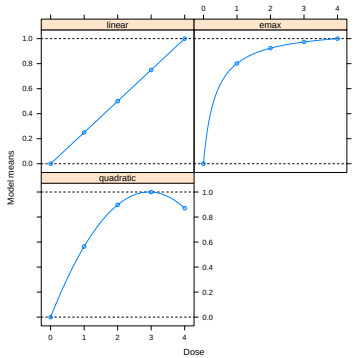
- In addition to PoC, also comparison of hightest dose vs placebo is statistically significant (H_3 is **rejected**).

Simulation study

- Parallel group design with $k = 4$ doses and placebo
- Total sample of $N = 100$ for the entire trial, split between the two stages with 60 in the first and 40 for the second stage
- Set of candidate models $\mathcal{M}_1$:



and



- Compare adaptive MCP-Mod with an adaptive combination test using Dunnett adjusted p-values and the inverse normal combination function

Simulation study: Interim selection rule

- At interim, select placebo d_0 and the "best" dose for the second stage
- Since the total sample size is fixed, we have an implicit sample reassessment as well
- In the final analysis, we are interested in testing H_i only for the selected dose i, but using totality of information from both stages and strongly controlling the overall Type I error rate

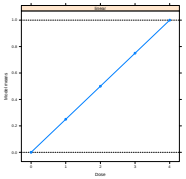
Adaptive MCP-Mod

- Use "best" model to select the dose with the largest response, which may not be the dose with the largest interim estimate

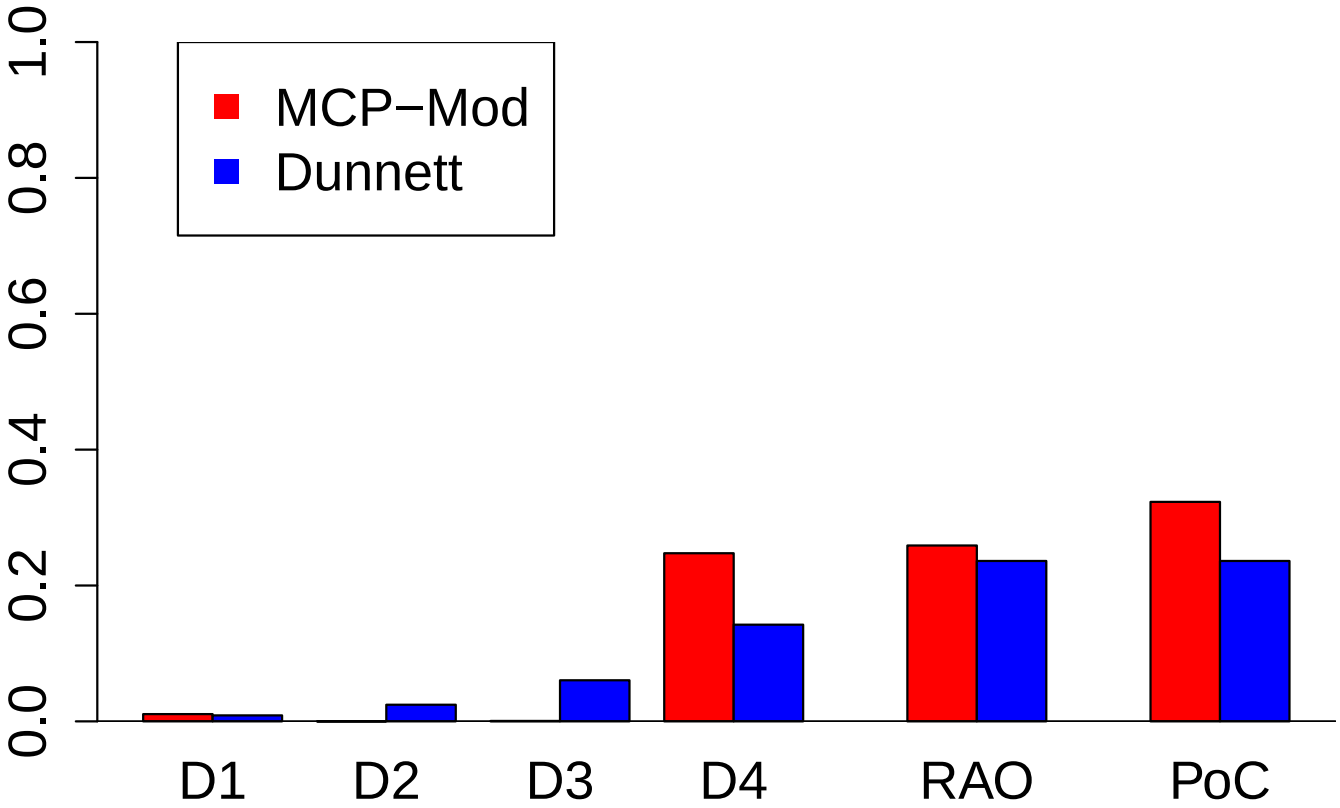
Adaptive design using Dunnett adjusted p-values

- Select the dose with largest interim estimate.

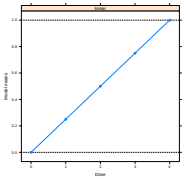
Simulation results



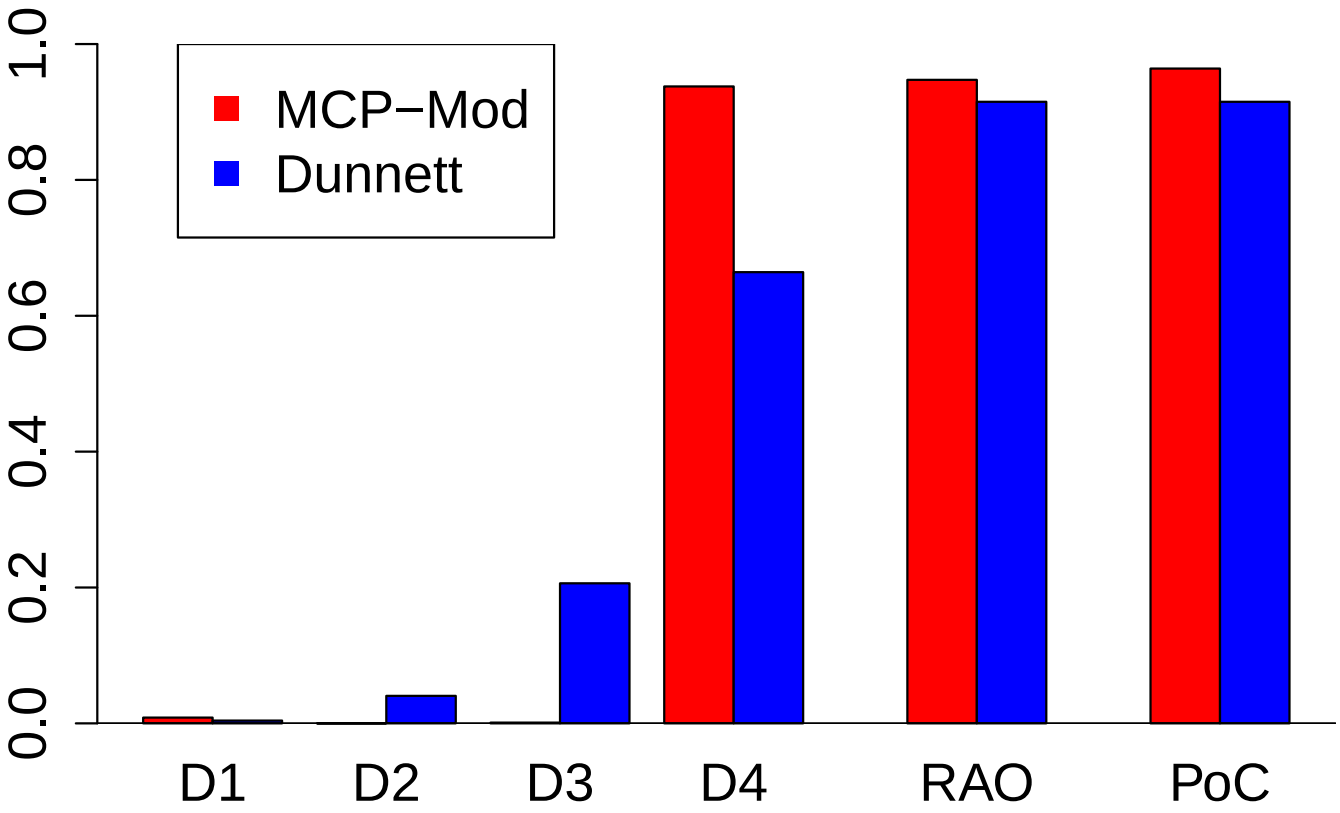
True linear model with $\mu_4 = 0.4\sigma$, 2 candidate models in $\mathcal{M}_1$



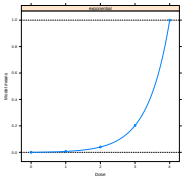
Simulation results



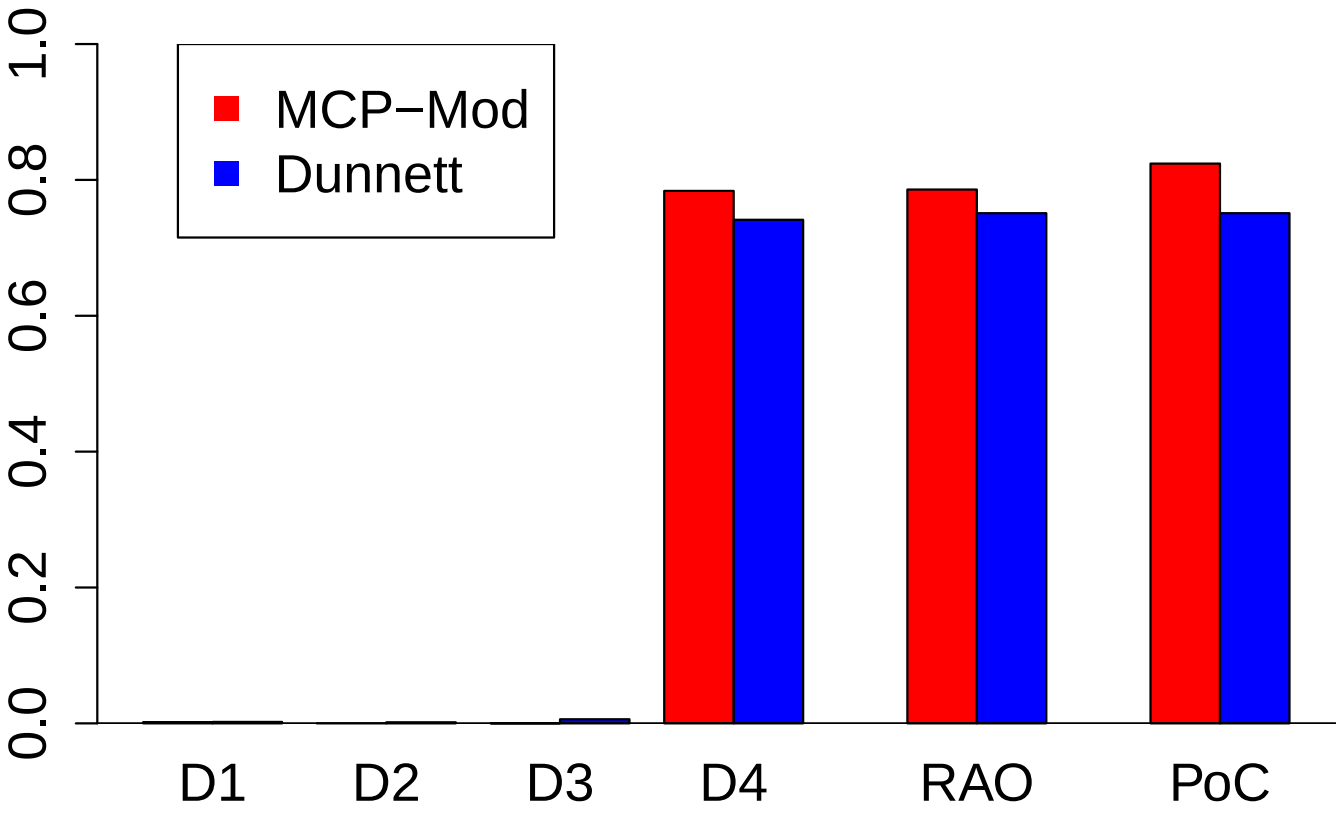
True linear model with $\mu_4 = \sigma$, 2 candidate models in $\mathcal{M}_1$



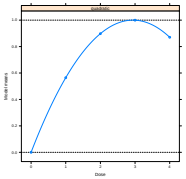
Simulation results



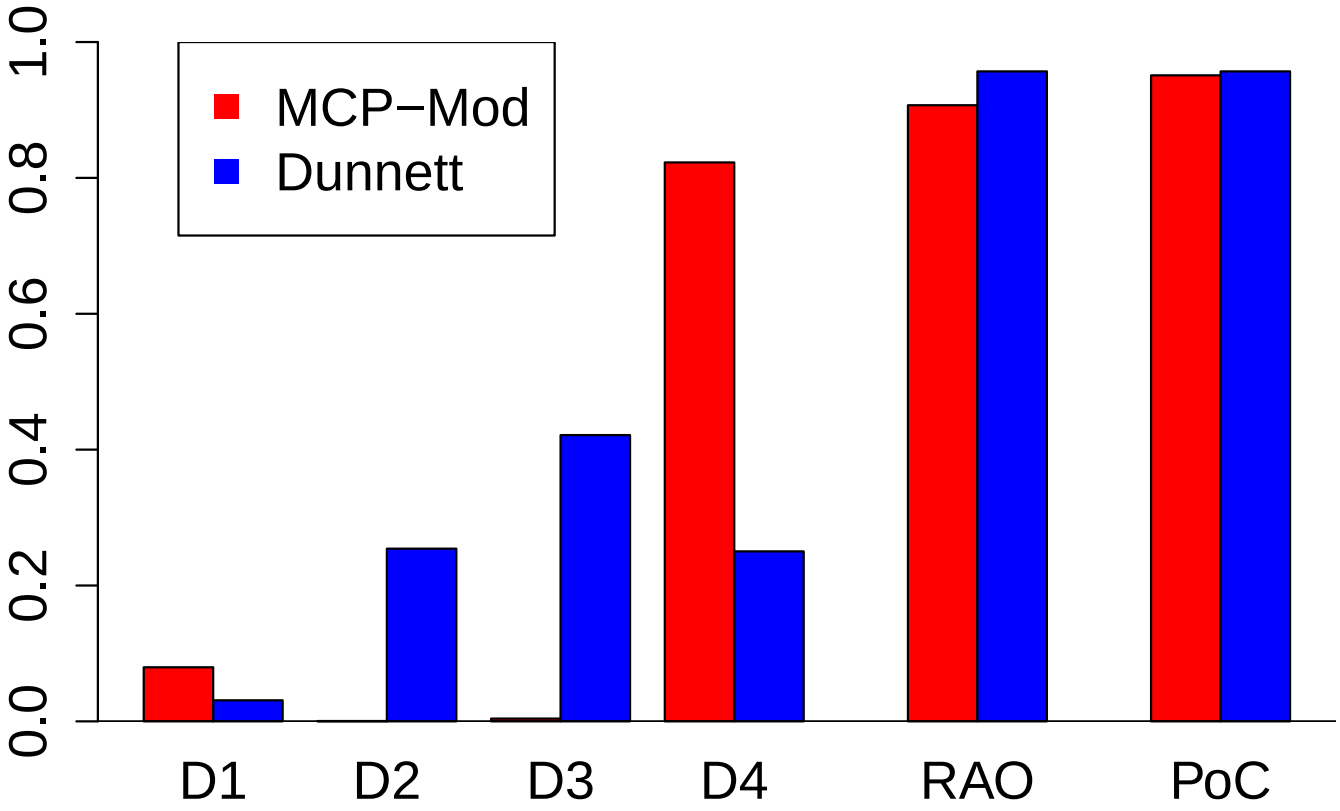
True exponential model with $\mu_4 = \sigma$, 2 candidate models in $\mathcal{M}_1$



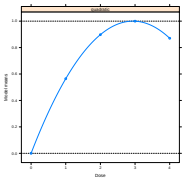
Simulation results



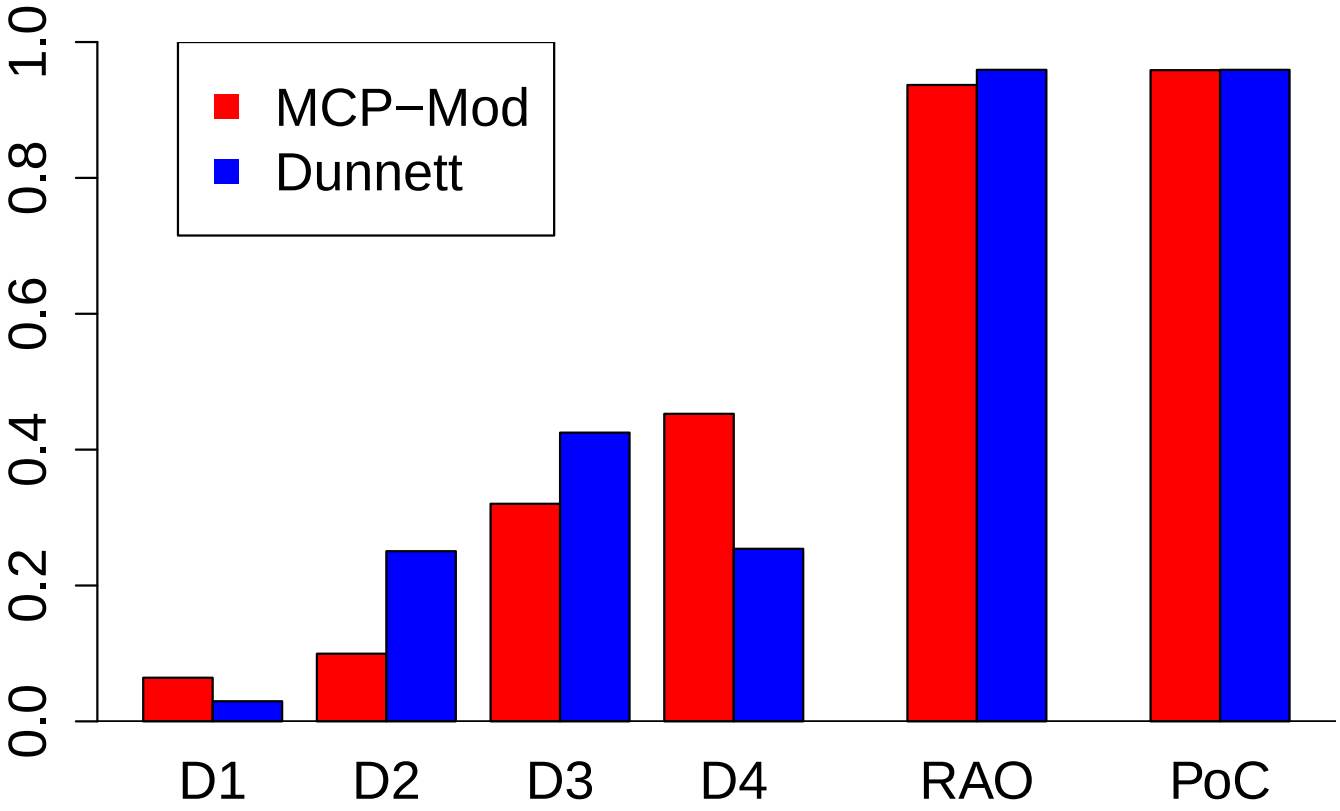
True quadratic model with peak at d_3 and $\mu_3 = \sigma$, 2 candidate models in $\mathcal{M}_1$



Simulation results



True quadratic model with peak at d_3 and $\mu_3 = \sigma$, 3 candidate models in $\mathcal{M}_1$



Summary

- Applying the closed test procedure to the original MCP-Mod approach, we obtain pairwise dose-control comparisons almost free of charge
- Unconstrained and constrained MCPMod versions available to ensure type I error control (not shown today)
- Using combination test principle enables adaptive interim analysis to change models, sample sizes, doses, ...
- "Optimal" design depends on which power is of main interest (PoC, RAO, specific dose, ...)
- More work to do: more advanced interim selection rules, sample size reassessment based on conditional power arguments, ...

References

- P. Bauer, F. Bretz, V. Dragalin, F. König, and G. Wassmer.
  Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls.
  Statistics in Medicine - Early View (2015)
  http://dx.doi.org/10.1002/sim.6472 (Free Download)
- Bauer, P. and Kieser, M. (1999). Combining different phases in the development of medical treatments within a single trial. Statistics in Medicine, 18:1833-1848.
- Bornkamp B, Pinheiro J, Bretz F (2014). R-Package “DoseFinding”
- Bretz F, König F, Brannath W, Glimm E, Posch M (2009). Adaptive Designs for Confirmatory Clinical Trials. Statistics In Medicine 2009 Apr 15;28(8):1181-217.
- Bretz, F., Pinheiro, J.C., and Branson, M. (2005) Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics, 61(3), 738-748.
- CHMP (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. EMA/CHMP/SAWP/757052/2013, available at www.ema.europa.eu
- Hommel, G. (2001). Adaptive modidcations of hypotheses after an interim analysis. Biometrical Journal, 43(5):581-589.
- König F, Bornkamp B, Bretz F, Glimm E (2014). Confirmatory testing for a beneficial treatment effect in dose-response studies using MCP-Mod. Presentation ISCB 2014, Vienna, Austria.
- Krasnozhon S, Graf A, Bornkamp B, Bretz F, Wassmer G, König F (2014). Adaptive designs for confirmatory model based decisions using MCP-Mod. Presentation ISCB 2014, Vienna, Austria.
- Marcus R, Peritz E, Gabriel KR (1976). On closed testing procedure with special reference to ordered analysis of variance. Biometrika 1976; 63:655-660.
- Pinheiro, J., Bornkamp, B., Glimm, E., and Bretz, F. (2014) Model-based dose finding under model uncertainty using general parametric models. Statistics in Medicine 33:1646-1661.

Type I error control using proposed MCP-Mod

- For $\mu_0 = \mu_1 = ... = \mu_k$ everything is fine
- (or if you use two-sided instead of one-sided tests)
- If in reality $\mu_i \leq \mu_0$ holds for some $i > 0$, a misspecification of the profile (with some negative weights c_i for $i > 0$) might lead to situations where this test rejects with probability $> \alpha$

- For strong control of the Type I error rate you have to either
  - assume $\mu_i \geq \mu_0$ for all $i > 0$, or
  - test only 2 doses against control (proof not shown here), or
  - allow only candidates models and/or doses ($i > 0$) with $c_{mi} \geq 0$ (very restrictive: might exclude to many model and doses), or
  - impose constraint $c_{mi} \geq 0$ for $i > 0$ when determining optimal contrasts for any candidate model m (only for placebo we allow $c_{m0} < 0$)

Constrained MCP-Mod

- Replace original, "unconstrained" optimal contrast coefficients by constrained coefficients satisfying $c_{mi} \geq 0$ for all $i = 1, \ldots, k$

Unconstrained MCP-Mod

Dose	linear	emax1	emax2	sigEmax	quadratic
0	-0.536	-0.861	-0.770	-0.515	-0.709
0.15	-0.341	0.199	-0.102	-0.453	-0.078
0.5	0.114	0.317	0.343	0.310	0.694
1	0.764	0.345	0.529	0.658	0.093

Constrained MCP-Mod

Dose	linear	emax1	emax2	sigEmax	quadratic
0	-0.707	-0.861	-0.809	-0.782	-0.738
0.15	0.000	0.199	0.000	0.000	0.000
0.5	0.000	0.317	0.311	0.187	0.671
1	0.707	0.345	0.498	0.595	0.067

Simulation study

- Parallel group design with placebo and $k = 3$ doses $0.15, 0.5, 1$
- Normal responses with variance $\sigma^2 = 1$.
- Testing one-sided hypotheses at overall level $\alpha = 0.025$.
- Sample size per group $n = 30$
- Candidate models $\mathcal{M}$: Linear, 2 Emax ($ED_{50} = 0.025, 0.2$), sigmoid Emax ($ED_{50} = 0.4, h = 3$), and quadratic (maximum effect at dose 0.6)
- Compare power of unconstrained with constrained MCP-Mod
  - Further comparisons with step-down Dunnett and fixed sequence test shown in the Appendix.

Simulation results

- Difference in power of unconstrained and constrained MCP-Mod to reject individual hypotheses H_i, under different true models
  - No apparent loss in power when using constrained MCP-Mod
  - However, constrained MCP-Mod is less powerful for dose response signal detection (not shown)