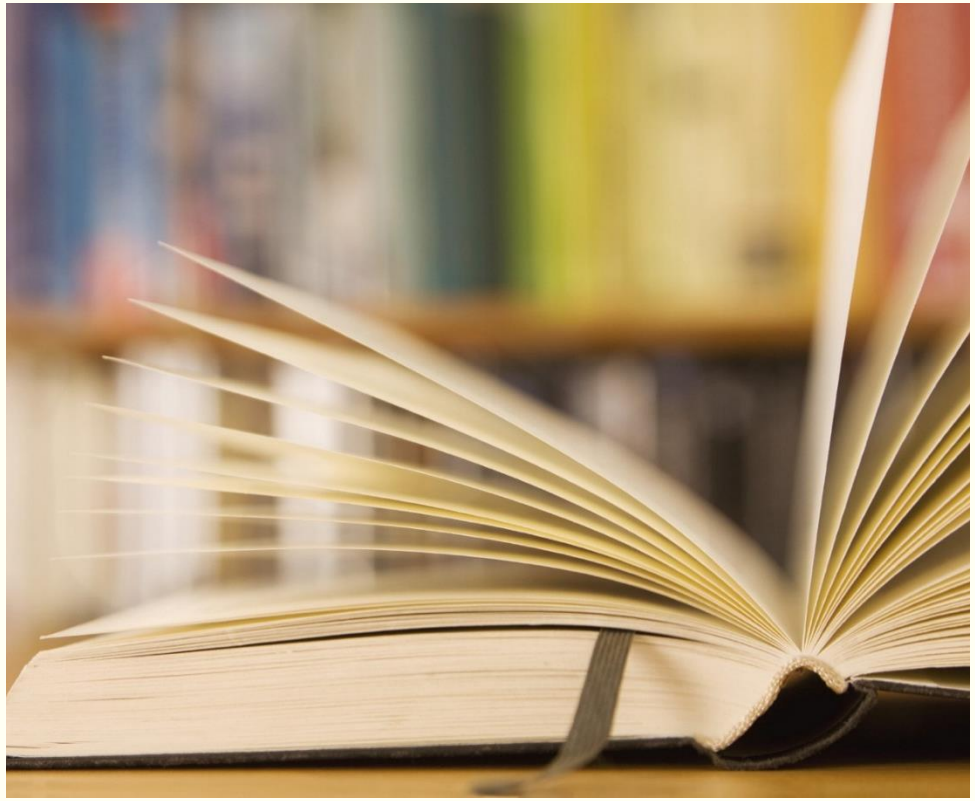
On the efficiency of the hybrid testing-modeling approach for binary responses

Min Yang (University of Illinois at Chicago)

Joint work with Andrew Yan (Evive Biotech)



A motivating dose-finding study

- Nonalcoholic steatohepatitis (NASH)
- 4 dose levels: 0mg, 0.3mg, 1mg, 3mg
- Liver fibrosis response (Y/N) – binary

- Primary objectives
  - Go/no go decision
  - If go, which dose(s)?

Phase II dose-response studies

- A bad decision could have expensive consequences
  - Pursue ineffective drug, terminate effective drug, recommend wrong dose level

- "Understanding the dose-response relationship and identifying the appropriate dose for Phase II clinical trials are probably the most critical, and yet most challenging, components of the clinical development program of a new therapy"
  - FDA (2016), Statistical review and evaluation qualification of statistical approach: MCP-MOD.

Our goal

- Efficient statistical methods for
  - study power
  - dose estimation

THE CURRENT APPROACHES

Classical approaches

- Multiple comparison procedure (MCP)
  - Dunnett's test, Cochran-Armitage trend test, or Williams' trend test

- Minimum effective dose (MED)
  - $MED = argmin_d(\hat{\pi}(d) \geq \hat{\pi}(d_1) + \delta, \hat{\pi}(d)^L \geq \hat{\pi}(d_1))$

- Pros – easy to implement & interpret, no assumptions about dose-response relationship
- But are they efficient?

Hybrid testing-modeling approach

- Parametric model
  - Pros: better predictive performance
  - Cons: model misspecification
- A set of models
  - M1, M2, M3,…
  - Cover a broad range of plausible dose-response shapes
- Control familywise type I error?

MCP-MOD (Pinheiro et al., 2014)

- Model assumption

$$g(\pi(d_i)) = \beta_{m0} + \beta_{m1} f_m^0(d_i)$$

- Optimal contrasts

$$c_m^{opt} = \hat{\boldsymbol{S}}^{-1} \left(\boldsymbol{\mu}_m^0 - \frac{\boldsymbol{\mu}_m^{0\prime} \hat{\boldsymbol{S}}^{-1} 1}{1' \hat{\boldsymbol{S}}^{-1} 1} \right)$$

- Multiple comparisons

$$z_m = \frac{(c_m^{opt})' \hat{\boldsymbol{\mu}}}{\sqrt{(c_m^{opt})' \hat{\boldsymbol{S}} c_m^{opt}}}, \; m = 1, \ldots, M.$$

- MED Estimation
  - Best model or weighted average

More about MCP-MOD

- Generality and fast computation
  - Continuous, binary, count, time-to-event endpoints,…
  - Fast computation

- Key assumption of MCP-MOD: multivariate normal distribution
  - Binary responses and small/moderate sample sizes?

Permutation test (Klingenberg, 2009)

$$T_s = (-1)^{I(\hat{\pi}_s(d_{max}) \leq \hat{\pi}_s(d_1))}\{-2[\log L(\boldsymbol{y}, \boldsymbol{n}; M_0) - \log L(\boldsymbol{y}, \boldsymbol{n}; M_s)] - 2df_s\}.\ m = 1, \ldots, M.$$

$$P_s^{obs} = \frac{1}{B}\sum_{b=1}^{B} I(T_s^{(b)} \geq T_s^{obs}),$$

$$P_s^{b} = \frac{1}{B}\sum_{l=1}^{B} I(T_s^{(l)} \geq T_s^{(b)}),$$

- model fit for each permuted data set
- Permutation distribution of minimum p-value

- Pros: normality assumption not required
- Cons: Likely to fail when nonlinear models are fitted

Design issues

- A uniform sample size allocation is often used.

- Is this a good choice?

Questions need to be addressed

- How to choose a set of candidate models?
- Efficient designs?
- Hybrid approach
  - POC: MCP-Mod or permutation test
  - Estimation: best model, average of all models, or average of significant models
- Overall recommendation

SET OF CANDIDATE MODELS

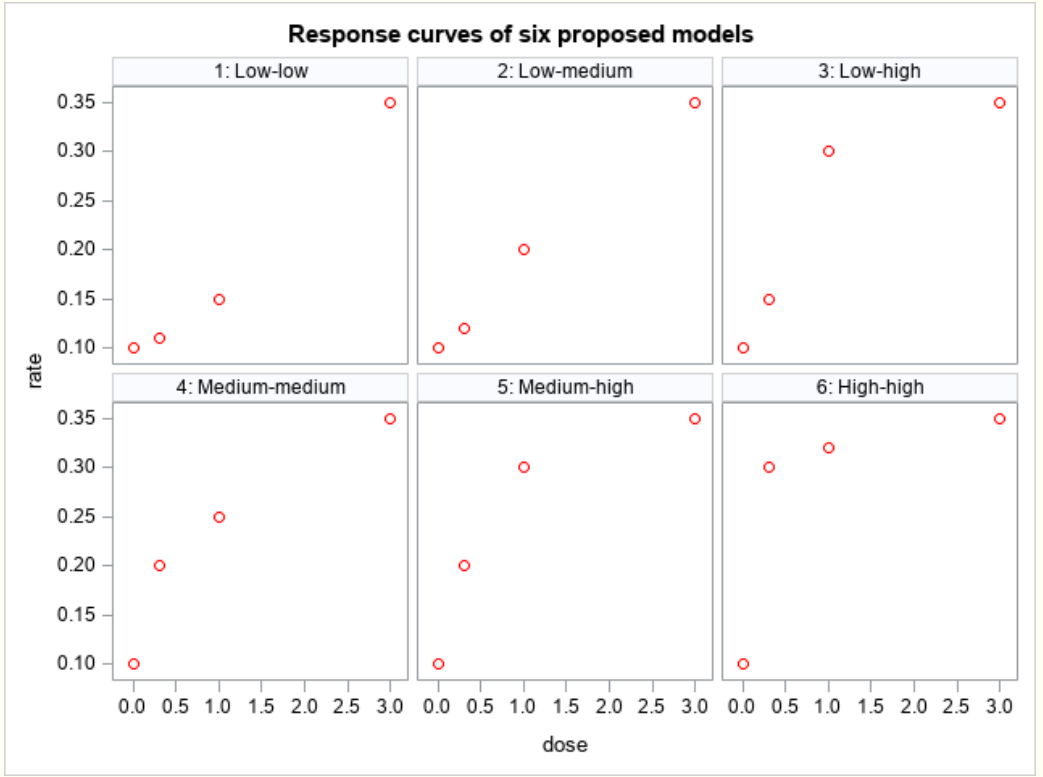
How to select a candidate set of models?

- Plausible dose-response models
- Stable parameter estimation
  - Permuted data
- Cover a broad range of dose-response curves
  - Likely nonlinear feature

The NASH study revisit

- Preliminary dose-scenario information
  - Monotone increasing

- Six scenarios for the two mid dose levels (0.3mg and 1mg)
  - Low-low
  - Low-medium
  - Low-high
  - Medium-medium
  - Medium-high
  - High-high



Plausible models

- 76 dose-ranging studies: Emax model (FDA)

$$\beta_0 + \frac{\beta_1 d}{d+\beta_2}, \quad \beta_1 > 0, \beta_2 > 0$$

- Reduced Emax Models:
  - $\beta_0 + \frac{\beta_1 d}{d+5.3}$: low-medium
  - $\beta_0 + \frac{\beta_1 d}{d+0.7}$: medium-medium
  - $\beta_0 + \frac{\beta_1 d}{d+0.4}$: medium-high
  - $\beta_0 + \frac{\beta_1 d}{d+0.15}$: high-high
- Not adequate for low-low and low-high scenarios
  - Convex and S-shape

A candidate set of models:

6 candidate Models:

$\beta_1 + \beta_2(e^{\frac{d}{1.5}} - 1)$: low-low

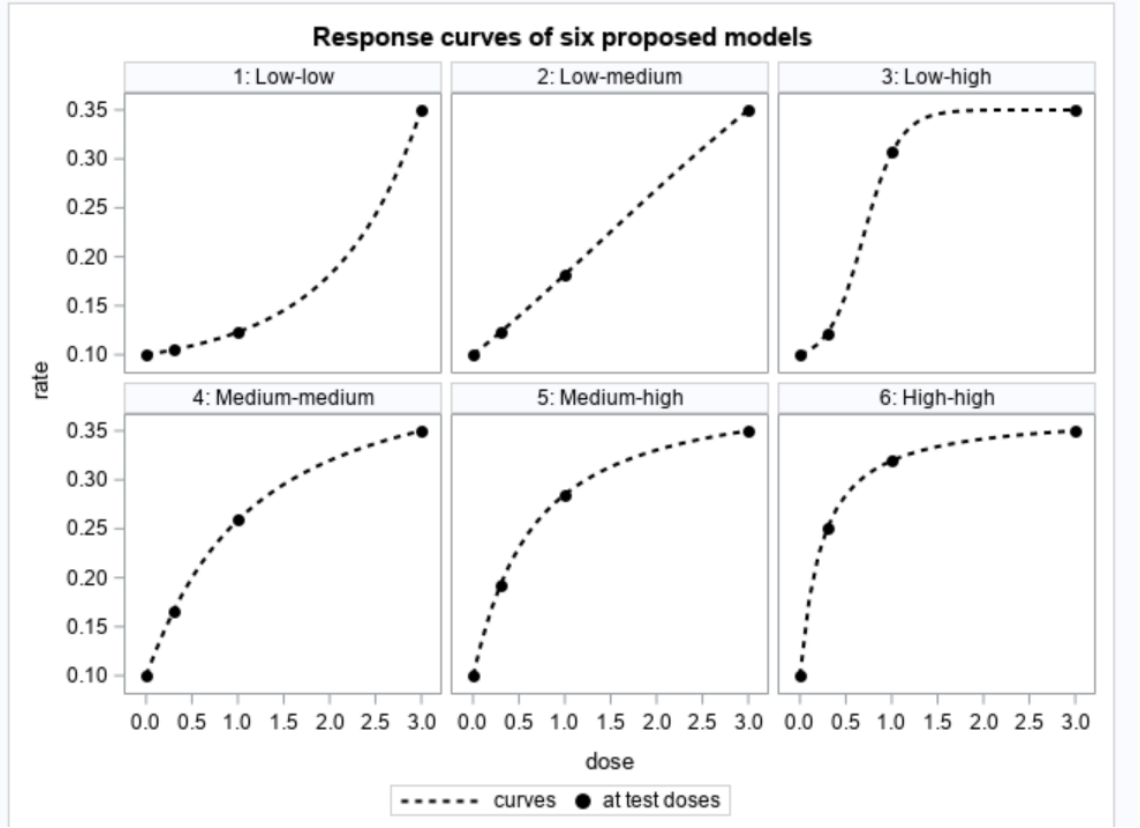
$\beta_1 + \frac{\beta_2 d}{d+5.3}$: low-medium

$\beta_1 + \frac{\beta_2}{1+e^{3-5d}}$:low-high

$\beta_1 + \frac{\beta_2 d}{d+0.7}$:medium-medium

$\beta_1 + \frac{\beta_2 d}{d+0.4}$:medium-high

$\beta_1 + \frac{\beta_2 d}{d+0.15}$:medium-high



ABOUT OPTIMAL DESIGNS

Toy example

Method I

$$Y_1 = A + \epsilon_1; \quad Y_2 = B + \epsilon_2;$$
$$Y_3 = C + \epsilon_1; \quad Y_4 = D + \epsilon_4.$$

$$\hat{A} = Y_1; \quad \hat{B} = Y_2;$$
$$\hat{C} = Y_3; \quad \hat{D} = Y_4.$$

$$Var(\hat{A}) = Var(\hat{B}) = Var(\hat{C}) = Var(\hat{D}) = \sigma^2$$

Hypothesized constrains

- Per use
  - $100K
  - One week

Method II

$$Y_1 = A + D + \epsilon_1; \quad Y_2 = A - D + \epsilon_2;$$
$$Y_3 = B + C + \epsilon_3; \quad Y_4 = B - C + \epsilon_4.$$

$$\hat{A} = (Y_1 + Y_2)/2; \quad \hat{B} = (Y_3 + Y_4)/2;$$
$$\hat{C} = (Y_3 - Y_4)/2; \quad \hat{D} = (Y_1 - Y_2)/2.$$

$$Var(\hat{A}) = Var(\hat{B}) = Var(\hat{C}) = Var(\hat{D}) = \sigma^2/2$$

Method III

$$Y_1 = A - B + C - D + \epsilon_1; \quad Y_2 = A + B - C + D + \epsilon_2;$$
$$Y_3 = A - B + C + D + \epsilon_3; \quad Y_4 = -A + B + C + D + \epsilon_4.$$

$$\hat{A} = (Y_1 + Y_2)/2; \quad \hat{B} = (Y_1 + Y_2 - Y_3 + Y_4)/2;$$
$$\hat{C} = (Y_1 + Y_4)/2; \quad \hat{D} = (-Y_1 + Y_3)/2.$$

$Var(\hat{A}) = Var(\hat{C}) = Var(\hat{D}) = \sigma^2/2$ and $Var(\hat{B}) = \sigma^2$

Method IV

$$Y_1 = A + B + C - D + \epsilon_1; \quad Y_2 = A + B - C + D + \epsilon_2;$$
$$Y_3 = A - B + C + D + \epsilon_3; \quad Y_4 = -A + B + C + D + \epsilon_4.$$

$$\hat{A} = (Y_1 + Y_2 + Y_3 - Y_4)/4; \quad \hat{B} = (Y_1 + Y_2 - Y_3 + Y_4)/4;$$
$$\hat{C} = (Y_1 - Y_2 + Y_3 + Y_4)/4; \quad \hat{D} = (-Y_1 + Y_2 + Y_3 + Y_4)/4.$$

$$Var(\hat{A}) = Var(\hat{B}) = Var(\hat{C}) = Var(\hat{D}) = \sigma^2/4$$

Comparison

Same cost ($400k) and length of time (4 weeks):

Method	I	II	IV
Variance	σ^2	$\frac{\sigma^2}{2}$	$\frac{\sigma^2}{4}$

To achieve the same variance ($\frac{\sigma^2}{4}$):

Method	I	II	IV
Cost	$1600k	$800k	$400k
Length of time	16 weeks	8 weeks	4 weeks

What do we learn from this toy example?

- A hybrid approach could be helpful
- An efficient design matters
- A hybrid approach + an efficient design could make a significant difference

Optimal designs for binary responses

- **Depend on model form and parameter values**
- **A dilemma**
  - **Catch-22?**
- **A gap between theory and practice**
  - **Looking backward**

Design an efficient clinical trial looking forward?

- "*It should be noted that both the model shape and model parameters need to be known a priori in order to design a dose-ranging study based on the D-optimal design. This is almost never the case in clinical drug development*."
  - FDA (2015)

- Can we design an efficient clinical trial looking forward?
  - when neither model form nor model parameters is known.

Theoretical result

$$g(\pi(d_i)) = \beta_{m0} + \beta_{m1} f_m^0(d_i)$$

Theorem 1. *For a clinical trial with binary responses, under Model (1) with logit link, suppose the dose response rate range is* $[r_1, r_2]$. *Let* ξ *be a design with half of observations at the dose level with response rate* r_1 *and half of observations at the dose level with response rate* r_2. *Then* ξ *is the D-optimal design for* (β_0, β_1) *regardless of parameter values if there exists* $x_2 \in (0, x^*)$ *such that* $[r_1, r_2]$ *is a subset of* $[\frac{e^{x_1}}{1+e^{x_1}}, \frac{e^{x_2}}{1+e^{x_2}}]$. *Here* $x^* > 0$ *is the solution to*

$$e^x + x + 1 - xe^x = 0 \tag{7}$$

and x_1 *is the solution to*

$$(1 - e^x)(x_2 - x) - 2(1 + e^x) = 0. \tag{8}$$

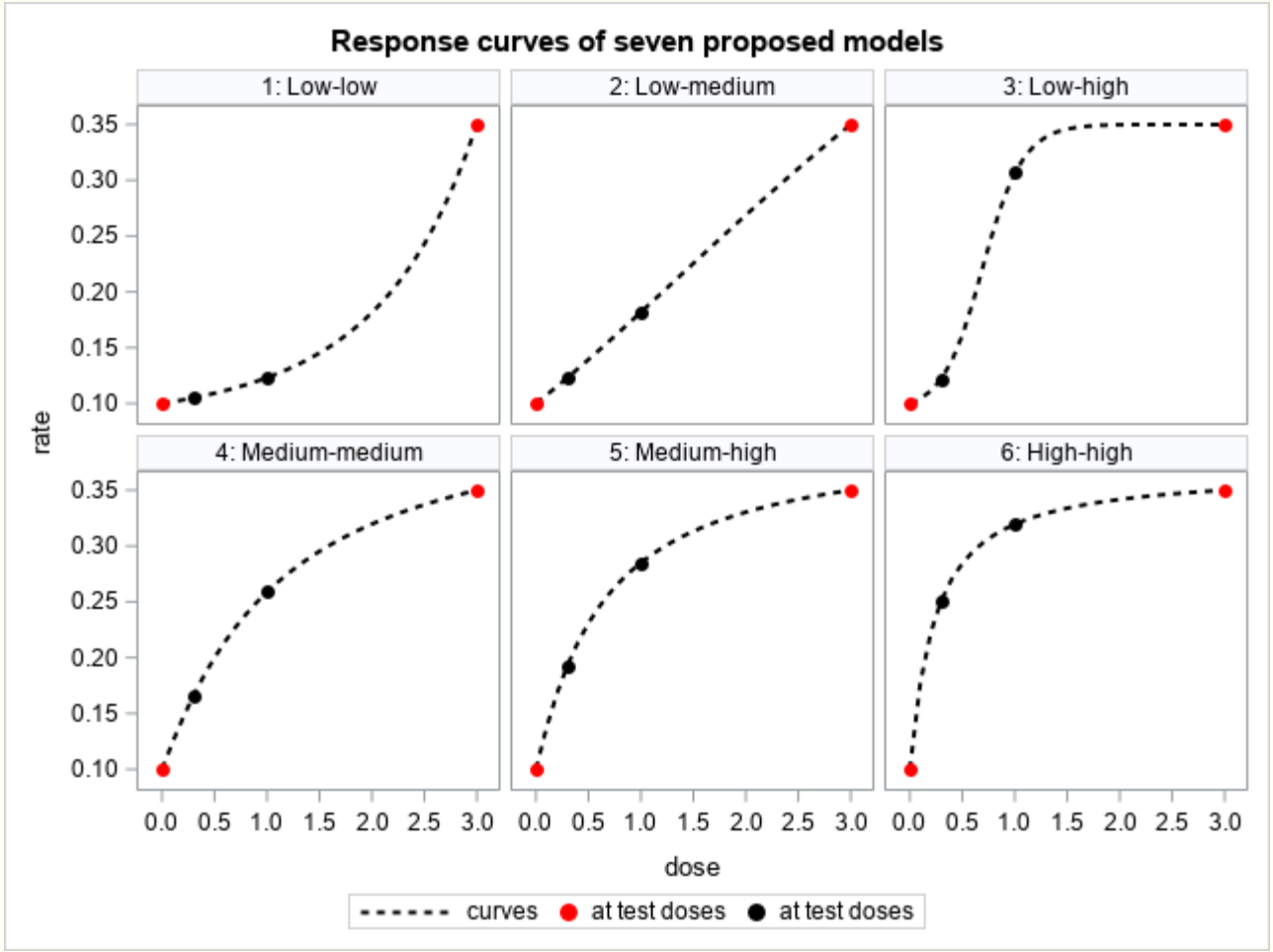
Application

Half of observations at dose with

- Minimum response rate
- Maximum response rate

The same optimal design

- Independent of model shape
- Independent of parameter values



Looking forward?

- Design an efficient clinical trial looking forward?

Yes

SIMULATION STUDIES

Design

- 1:1:1:1 (30:30:30:30)
- 2:1:1:2 (40:20:20:40)
- 3:1:1:3 (45:15:15:45)

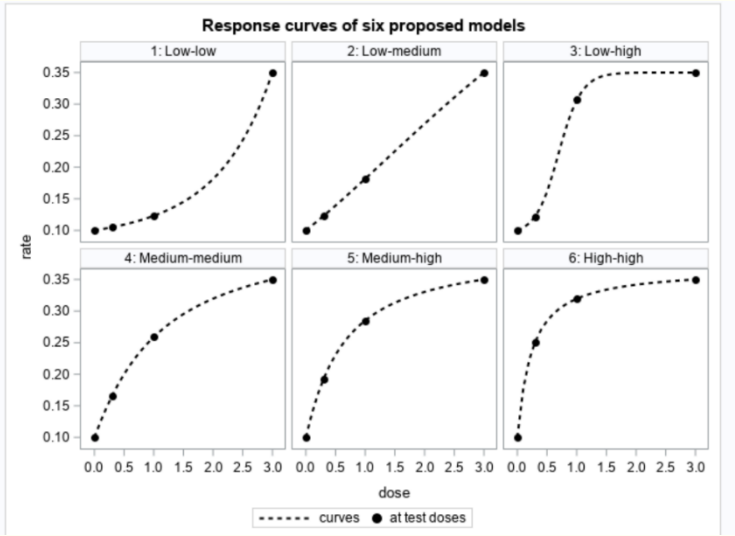
Analysis: competing methods

- Classical
  - CA
  - Dunnet
  - Williams

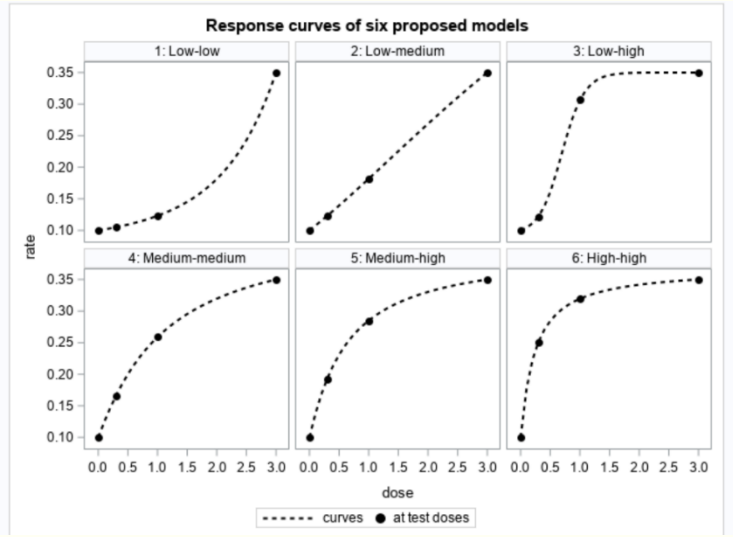
- Hybrid
  - MCP-MOD
    - 6 models
    - 7 models
  - Permutation testing-modeling
    - 6 models
    - 7 models

Two sets of candidate models

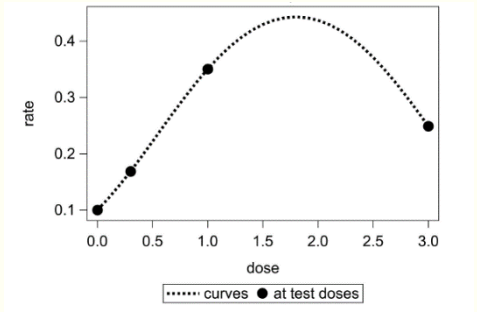
- 6 models



- 7 models



+



Dose estimation methods

- Best fitting model
- Weighted average
  - All models
  - Significant models only

Assessments

- Familywise type I error rate
- Study power
- MED estimation

Familywise type I error rate

- Dose scenarios
  - 0.1, 0.1, 0.1, 0.1
- 1000 simulations
- 5% level of significance (1-sided)

Familywise type I error rate comparison

Type I error rate

Sample Size Plan	PTM		MCP-Mod		CA	Dunnet	Williams
	6 model set	7 model set	6 model set	7 model set			
I	0.044	0.047	0.020	0.011	0.047	0.054	0.043
II	0.062	0.058	0.014	0.010	0.054	0.046	0.051
III	0.050	0.048	0.029	0.020	0.050	0.057	0.058

MCP-MOD tends to be conservative

All other methods control type I error rate well

Dose-response scenarios for assessing power

- 3 parametric model groups

Group I	Group II	Group III
M1: $\beta_0 + \beta_1[\exp(d/1.5) - 1]$	M8: $\beta_0 + \beta_1[\exp(d/2) - 1]$	M15: $\beta_0 + \beta_1 d$
M2: $\beta_0 + \frac{\beta_1 d}{5.3+d}$	M9: $\beta_0 + \frac{\beta_1 d}{4+d}$	M16: $\beta_0 + \beta_1\sqrt{d}$
M3: $\beta_0 + \beta_1/(1 + \exp(3 - 5d))$	M10: $\beta_0 + \beta_1/(1 + \exp(4 - 4d))$	M17: $\beta_0 + \beta_1 \log(d + 1)$
M4: $\beta_0 + \frac{\beta_1 d}{0.7+d}$	M11: $\beta_0 + \frac{\beta_1 d}{1.2+d}$	M18: $\beta_0 + \beta_1 \frac{1}{\sqrt{d+1}}$
M5: $\beta_0 + \frac{\beta_1 d}{0.4+d}$	M12: $\beta_0 + \frac{\beta_1 d}{0.5+d}$	M19: $\beta_0 + \beta_1 \frac{1}{d+1}$
M6: $\beta_0 + \frac{\beta_1 d}{0.15+d}$	M13: $\beta_0 + \frac{\beta_1 d}{0.2+d}$	M20: $\beta_0 + \beta_1 \exp(\exp(\frac{d}{\max d}))$
M7: $\beta_0 + \beta_1(-3.6d + d^2)$	M14: $\beta_0 + \beta_1(-4d + d^2)$	M21: $\beta_0 + \beta_1(-1.3\sqrt{d} + d)$

- 1 non-model-based group

Scenario I	Scenario II	Scenario III
(0.10, 0.12, 0.15, 0.35)	(0.10, 0.12, 0.2, 0.35)	(0.10, 0.12, 0.3, 0.35)
Scenario IV	Scenario V	Scenario VI
(0.10, 0.20, 0.25, 0.35)	(0.10, 0.20, 0.30, 0.35)	(0.10, 0.30, 0.32, 0.35)
	Scenario VII	
	(0.10, 0.30, 0.35, 0.20)	

Simulation set-ups for assessing power

- Total 28 different dose-response scenarios
  - each combined with 3 sample size allocations
  - 84 cases
- 500 simulations
- 5% level of significance (1-sided)

Table 1: POC power under non-model-based scenarios

Scenario	Sample size plan	6-model		7-model		CA	Dunnet	Williams
		PTM	MCP-Mod	PTM	MCP-Mod			
I	I	0.810	0.780	0.800	0.756	0.876	0.686	0.762
	II	0.896	0.872	0.884	0.806	0.924	0.794	0.868
	III	0.930	0.904	0.916	0.888	0.932	0.810	0.892
II	I	0.852	0.786	0.836	0.748	0.856	0.686	0.780
	II	0.878	0.852	0.870	0.830	0.898	0.748	0.848
	III	0.926	0.890	0.914	0.874	0.930	0.820	0.906
III	I	0.878	0.816	0.870	0.826	0.828	0.726	0.830
	II	0.916	0.886	0.908	0.882	0.888	0.828	0.904
	III	0.900	0.880	0.904	0.872	0.898	0.828	0.888
IV	I	0.772	0.706	0.768	0.684	0.716	0.712	0.788
	II	0.870	0.818	0.856	0.808	0.808	0.800	0.878
	III	0.902	0.876	0.900	0.866	0.876	0.830	0.902
V	I	0.818	0.736	0.824	0.720	0.726	0.720	0.812
	II	0.918	0.866	0.914	0.858	0.828	0.826	0.924
	III	0.930	0.906	0.922	0.896	0.854	0.840	0.930
VI	I	0.768	0.578	0.744	0.538	0.456	0.704	0.816
	II	0.890	0.822	0.876	0.806	0.716	0.852	0.924
	III	0.896	0.864	0.890	0.854	0.768	0.852	0.922
VII	I	0.478	0.292	0.508	0.334	0.036	0.702	0.642
	II	0.538	0.390	0.578	0.434	0.092	0.716	0.652
	III	0.496	0.420	0.544	0.462	0.150	0.652	0.610

Summary of power comparisons

Scenario	PTM		MCP-Mod		CA	Dunnet	Williams
	6-model set	7-model set	6-model set	7-model set			
Model-based Group I	-0.01	-	-0.05	-0.05	-0.08	-0.09	-0.02
Model-based Group II	0.01	-	-0.04	-0.05	-0.02	-0.10	-0.02
Model-based Group III	0.01	-	-0.03	-0.05	-0.01	-0.10	-0.02
Non-model-based	0.00	-	-0.06	-0.07	-0.10	-0.05	0.01

* PTM with 7 models as the reference

Power comparisons

- PTM outperforms MCP-MOD uniformly

- PTM outperforms other competing methods for most cases

Recommendation: PTM

Power difference between two candidate sets

- PTM
  - 7 models – 6 models

Power comparison between the 7-model set and 6-model set

Scenario	Mean	Min	Median	90% Quantile	Max
Increasing	-0.008	-0.03	-0.008	0.000	0.026
Downturn	0.033	0.004	0.025	0.070	0.076

Recommendation: 7 models

Power difference: 3 sample size allocations

- PTM
  - II – I
  - III – I

Scenario	Sample size plans	Mean	Min	Median	Max
Increasing	II vs I	0.079	0.034	0.081	0.132
	III vs I	0.103	0.034	0.100	0.152
Downturn	II vs I	0.041	-0.004	0.046	0.076
	III vs I	0.028	-0.070	0.041	0.098

Plan I: 30:30:30:30; Plan II: 40:20:20:40; Plan III: 45:15:15:45

Recommendation: III or II

MED Estimation

- Two types of MEDs
  - Unrestricted: any dosage within the dose range
  - Restricted: one of the test doses

- Three competing methods:
  - Best model
  - Weighted average of significant models only
  - Weighted average of all models

$$\widehat{MED}_s = \text{argmin}_{d\in(d_1,d_k]}\{\hat{\pi}_s(d) > \hat{\pi}_s(d_1) + \delta, \hat{\pi}_s^L(d) > \hat{\pi}_s(d_1)\}$$

$$\widehat{wMED} = \sum_{s\in\mathcal{M}_1} w_s\widehat{MED}_s / \sum_{s:\mathcal{M}_1} w_s$$

Setups for assessing MED estimation

- Same dose-response scenarios
  - 84 cases
  - 1000 simulations for each combination
  - MED estimation only if POC is positive

- MSE for unrestricted MED
  - For model-based dose-response scenarios only

- # of times restricted MED being correctly identified

Estimation methods comparison

Ratio of number of times the restricted MED is correctly identified

Estimation methods	Models	Mean	Min	Median	Max
Ave-sgnf/Ave-all	6-model	1.000	1.000	1.000	1.014
	7-model	0.990	0.961	0.992	1.017
Best/Ave-all	6-model	0.985	0.704	0.927	1.733
	7-model	0.910	0.677	0.887	1.288

Ratio of MSEs of the unrestricted MED

Estimation methods	Models	Mean	Min	Median	Max
Ave-sgnf/Ave-all	6-model	1.000	0.994	1.000	1.000
	7-model	0.999	0.967	0.992	1.017
Best/Ave-all	6-model	1.380	0.655	1.375	2.341
	7-model	1.313	0.614	1.313	2.347

- Recommendation: weighted average for all models

Set of candidate models comparison

Ratio of number of times the restricted MED is correctly identified (6-model set/7-model set)

Scenario	Mean	Min	Median	Max
Increasing	0.949	0.606	0.964	1.050
Downturn	0.617	0.295	0.704	0.881

- Recommendation: 7 models

Sample size plans comparison

Ratio of number of times the restricted MED is correctly identified

Scenario	Sample size plans	Mean	Min	Median	Max
Increasing	II/I	1.050	0.891	1.047	1.162
	III/I	1.086	0.968	1.077	1.311
Downturn	II/I	1.021	1.000	1.023	1.036
	III/I	0.967	0.944	0.948	1.029

Plan I: 30:30:30:30; Plan II: 40:20:20:40; Plan III: 45:15:15:45

- Recommendation: III or II

Overall recommendation

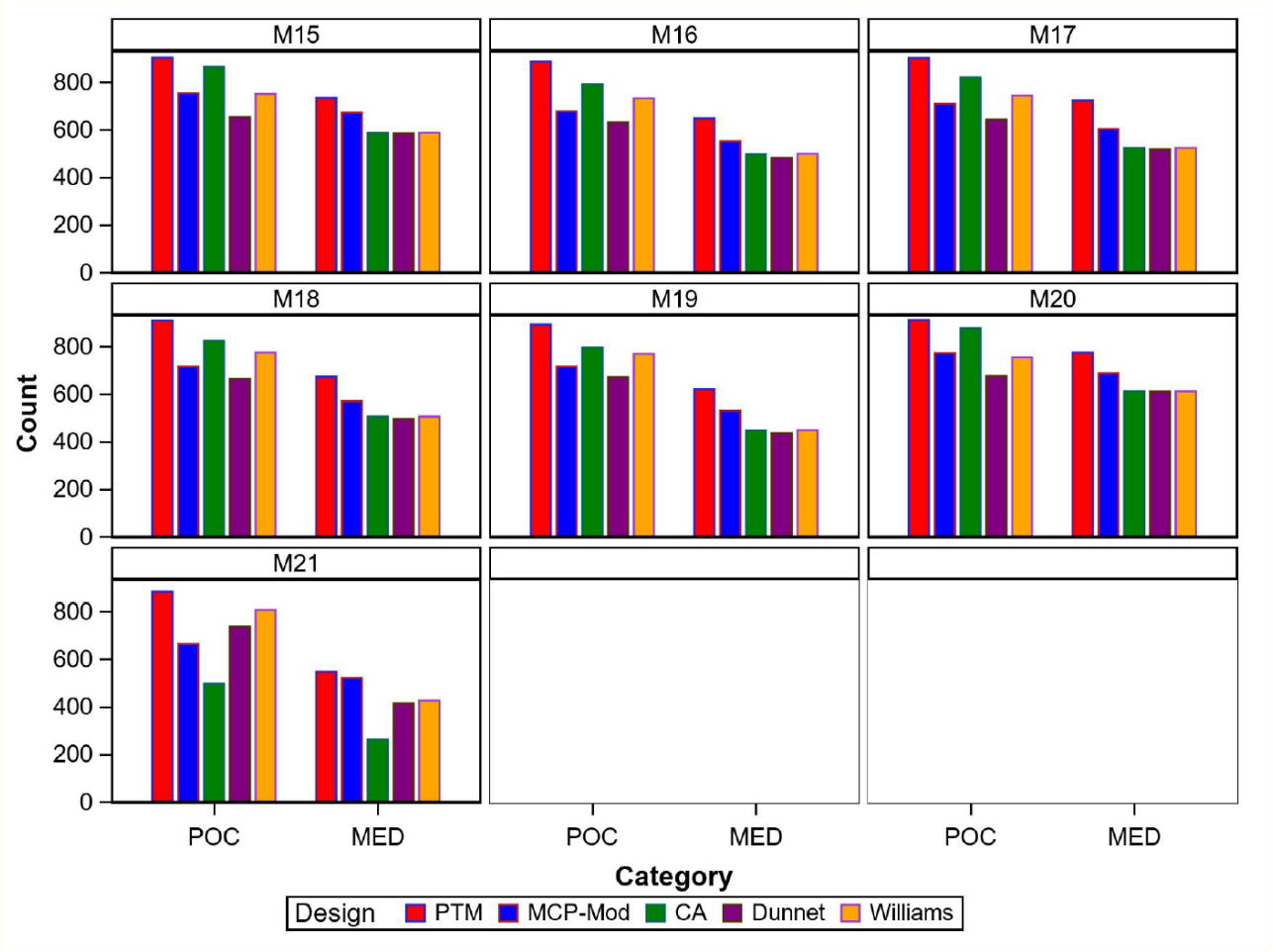
- Design
  - 3:1:1:3
- POC decision
  - PTM with 7-model set
- Dose estimation
  - Weighted average for all models

Direct comparison

- Proposed method compared with
  - Design: Uniform sample size allocation
  - Analysis method:
    - MCP-MOD with 7-model set
    - CA, Dunnet, Williams

- Dose-response scenario: parametric models
  - Group III (different model forms from the 7 models set)

Same total sample size (N=120)



Sample size comparisons

*Minimum total sample size needed to achieve the same efficiency**

Designs	M15	M16	M17	M18	M19	M20	M21
MCP-Mod (Uniform)	172	160	156	156	156	188	140
CA (Uniform)	308	236	276	264	272	360	352
Dunnet (Uniform)	308	236	276	264	272	360	260
William (Uniform)	308	236	276	264	272	360	260

* compared with PTM with a 3:1:1:3 sample size allocation ($N = 120$)

- Criterion: the percentage of correctly identified restricted MED
- 30000 repetitions
- $\hat{p}_{ptm} - \hat{p}_N > 0.01$ (the lower bound of 99% C.I. of $p_{ptm} - p_N$ >0)

AN EXAMPLE

Merck dose ranging trial revisit

- A pilot study for a new rheumatoid arthritis drug (Biedermann et al., 2006, 2007)
  - 120 patients equally allocated to a placebo (dose 0) and a high dose (dose 50).
  - The response rates were 35% at the placebo and 65% at the high dose.
  - Logit model $\pi(d_i) = \frac{e^{\beta(d_i-\alpha)}}{1+e^{\beta(d_i-\alpha)}},$ was fitted to those data
  - MLE of α and β is 25 and 0.025, respectively

- Dose ranging trial was to obtain MED (20% higher than placebo rate)

Set up

- 3 dose scenarios (based on the logistic model)

Three dose scenarios

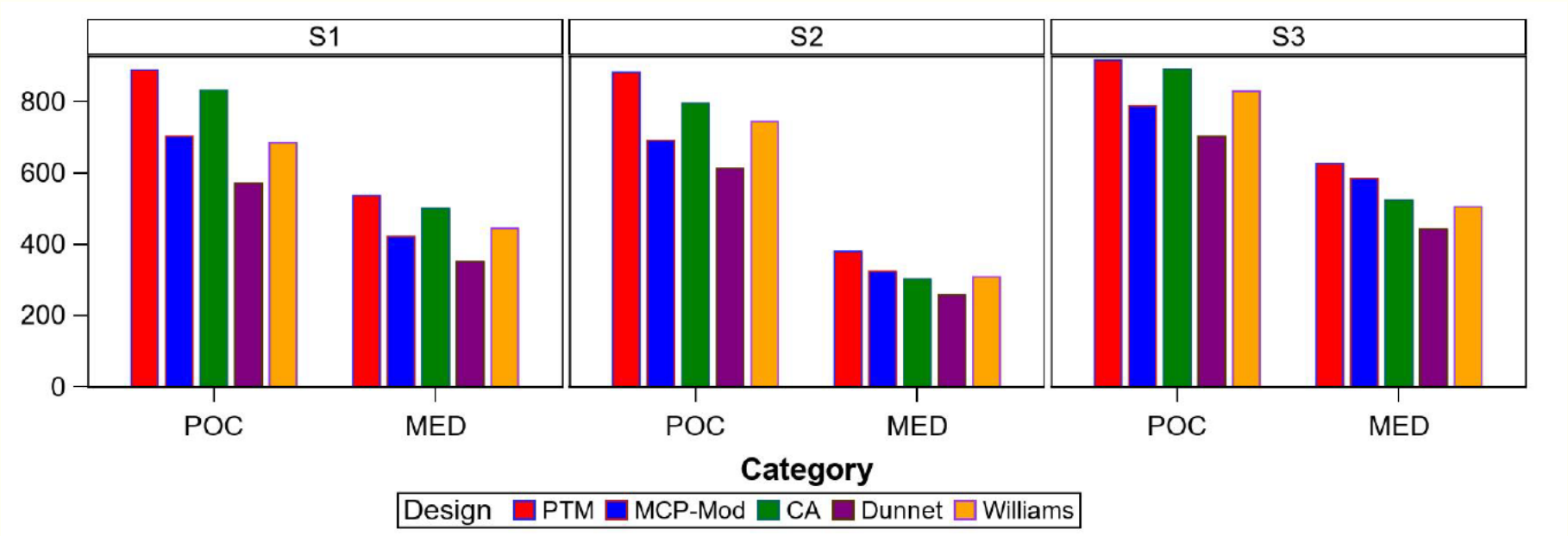
Scenario	S1	S2	S3
Dose levels	(0, 5, 20, 50)	(0, 17, 33, 50)	(0, 15, 45, 50)
Dose-response rates	(0.35, 0.38, 0.47, 0.65)	(0.35, 0.45, 0.55, 0.65)	(0.35, 0.44, 0.62, 0.65)

- 7 models (same models adjusted by the dose scale)

The seven candidate models

M1: $\beta_0 + \beta_1[\exp(d/25) - 1]$	M2: $\beta_0 + \frac{\beta_1 d}{88+d}$	M3: $\beta_0 + \beta_1/(1 + \exp(3 - 83d))$
M4: $\beta_0 + \frac{\beta_1 d}{11.7+d}$	M5: $\beta_0 + \frac{\beta_1 d}{6.7+d}$	M6: $\beta_0 + \frac{\beta_1 d}{2.5+d}$
M7: $\beta_0 + \beta_1(d^2 - 60d)$		

Same total sample size (N=120)



Sample size comparisons

*Minimum total sample size needed to achieve the same efficiency in the real example**

Designs	S1	S2	S3
MCP-Mod (Uniform)	144	240	136
CA (Uniform)	136	280	200
Dunnet (Uniform)	160	292	220
William (Uniform)	140	280	200

* compared with PTM with a 3:1:1:3 sample size allocation ($N = 120$)

- Criterion: the percentage of correctly identified restricted MED
- 30000 repetitions
- $\hat{p}_{ptm} - \hat{p}_N > 0.01$ (the lower bound of 99% C.I. of $p_{ptm} - p_N$ >0)

SUMMARY

Hybrid vs classical approaches

- Hybrid testing-modeling

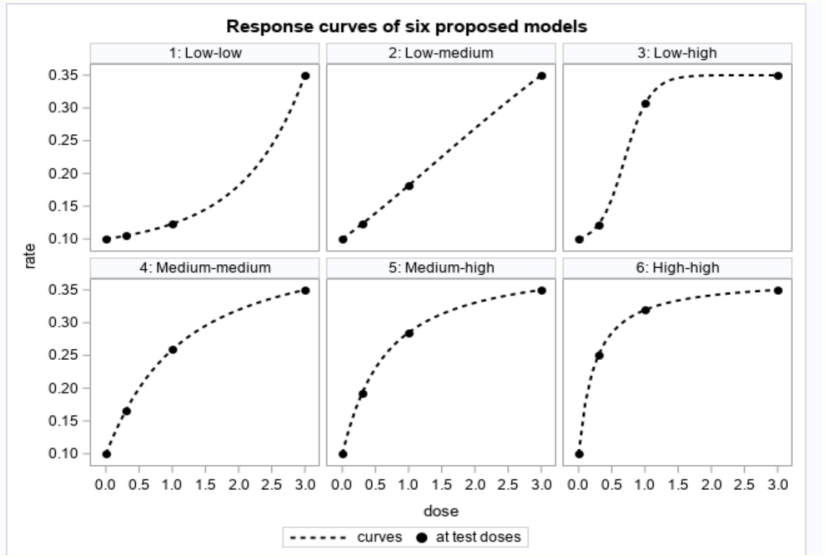
Which hybrid testing-modeling approach?

- PTM outperforms MCP-MOD uniformly
- PTM needs more computation resources
  - 3-4 vs 0.01 seconds per trial

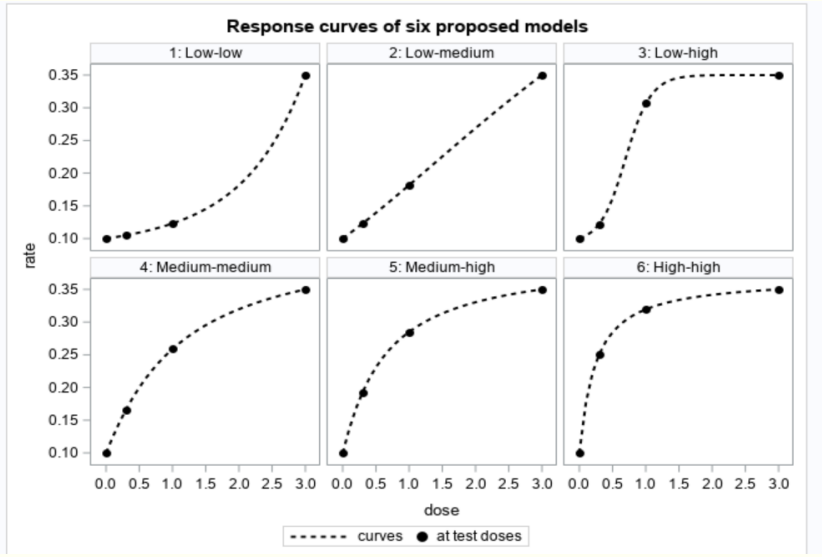
- PTM

Choice of candidate models

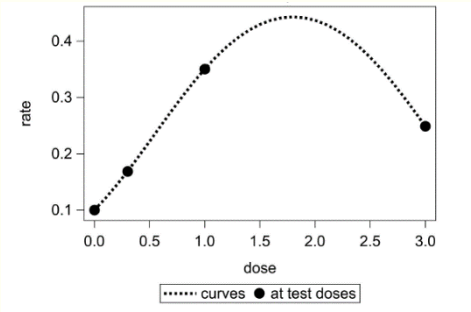
- 6 models



- 7 models

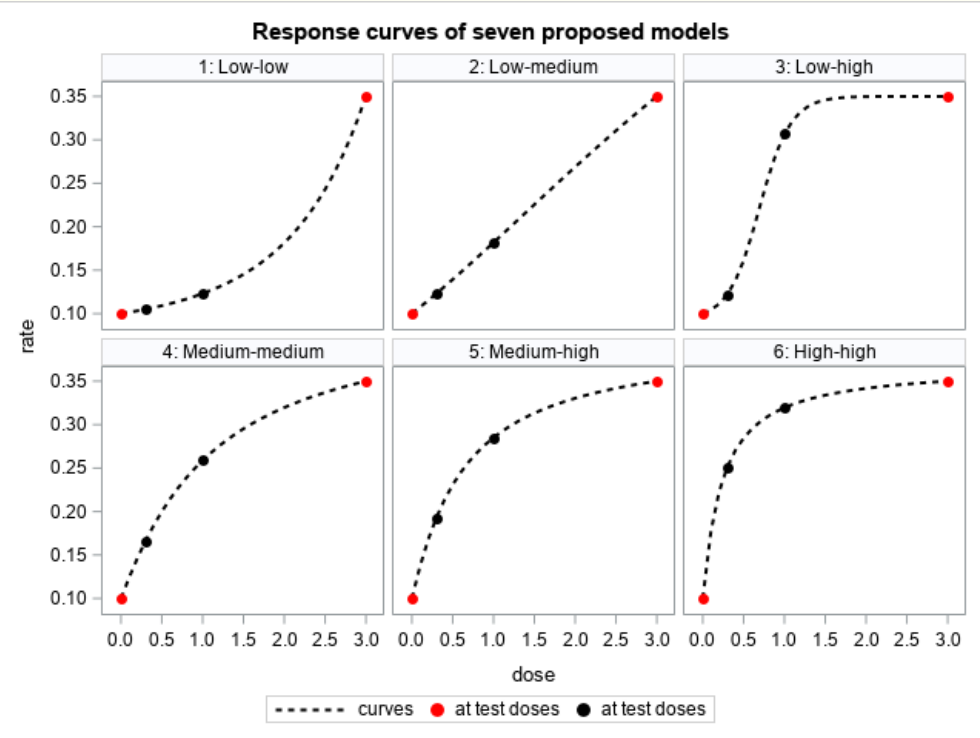


+



- Covers a broad range of plausible dose-response shapes

Design an efficient clinical study looking forward?



- Yes

Design?

- A Hybrid approach + an efficient design

How much is the gain?

*Minimum total sample size needed to achieve the same efficiency**

Designs	M15	M16	M17	M18	M19	M20	M21
MCP-Mod (Uniform)	172	160	156	156	156	188	140
CA (Uniform)	308	236	276	264	272	360	352
Dunnet (Uniform)	308	236	276	264	272	360	260
William (Uniform)	308	236	276	264	272	360	260

* compared with PTM with a 3:1:1:3 sample size allocation ($N = 120$)

- Significant

Thank you!