BMA-Mod : A Bayesian Model Averaging Strategy for Determining Dose-Response Relationships in the Presence of Model Uncertainty

A. Lawrence Gould
KOL
20 October, 2017

MERCK
INVENTING FOR LIFE

Overview and Motivation

Dose-Response Trials Overview

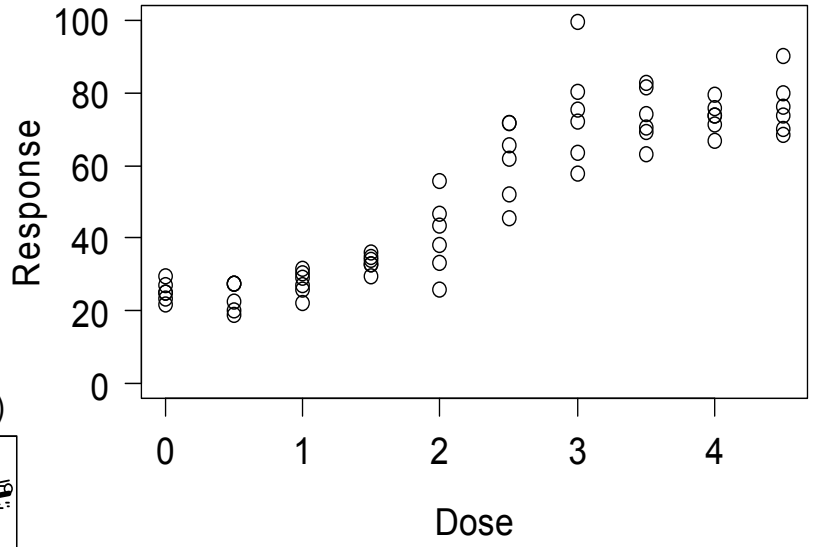
- Key objective: Find a 'correct' dose to carry forward
  - Applies at all pre-phase 3 stages
- Key issues to address (Ruberg, JBS1995)
  - Responses related to doses? (Proof of Concept) — Most Important
  - Dose(s) to carry forward (Target dose selection) — Most Important
  - Doses producing responses differing from control — Depends on trial design
  - Form of the dose-response relationship — Not useful for mixture models
- Extensive literature on determining dose-response relationships focuses on modeling or multiple comparisons

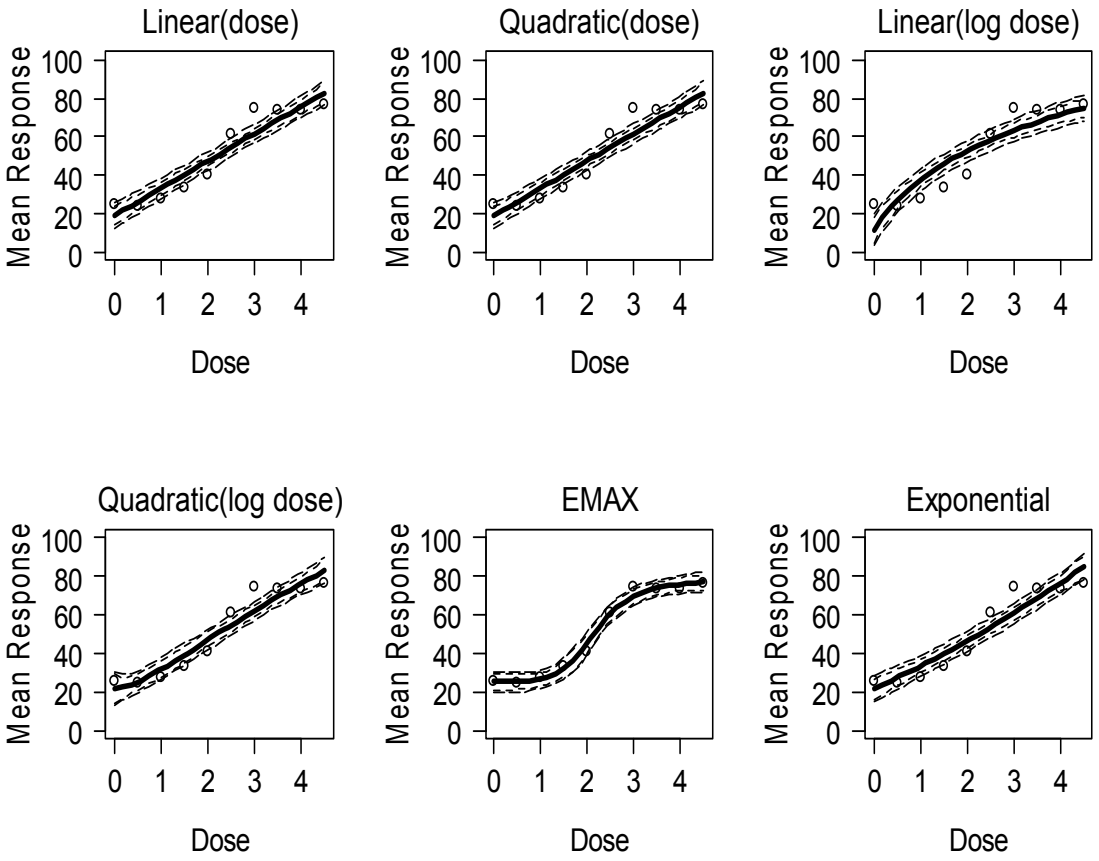
Motivation and Objective

- The model in a model-based approach could be wrong
- MCPMod: hypothesis-testing paradigm for including multiple models
  - Determine doses to carry forward by selecting most promising model, or by model averaging
  - Use nonlinear regression parameter estimation, multiple comparisons and asymptotic normality assumptions
- Objective of this presentation
  - Bayesian framework for carrying out multiple model analyses using estimation paradigm
    - Based on actual likelihoods
    - Directly address important issues
    - Definitive graphical displays
- Substantial literature on issues and examples related to optimal model selection and model weighting using Bayesian methods

A simple example: Obvious choice of model

- Data from Ruberg (JBS1995)
- Fitting a selection of models
  $\Rightarrow$ EMAX gives best fit
- Dots are observed means

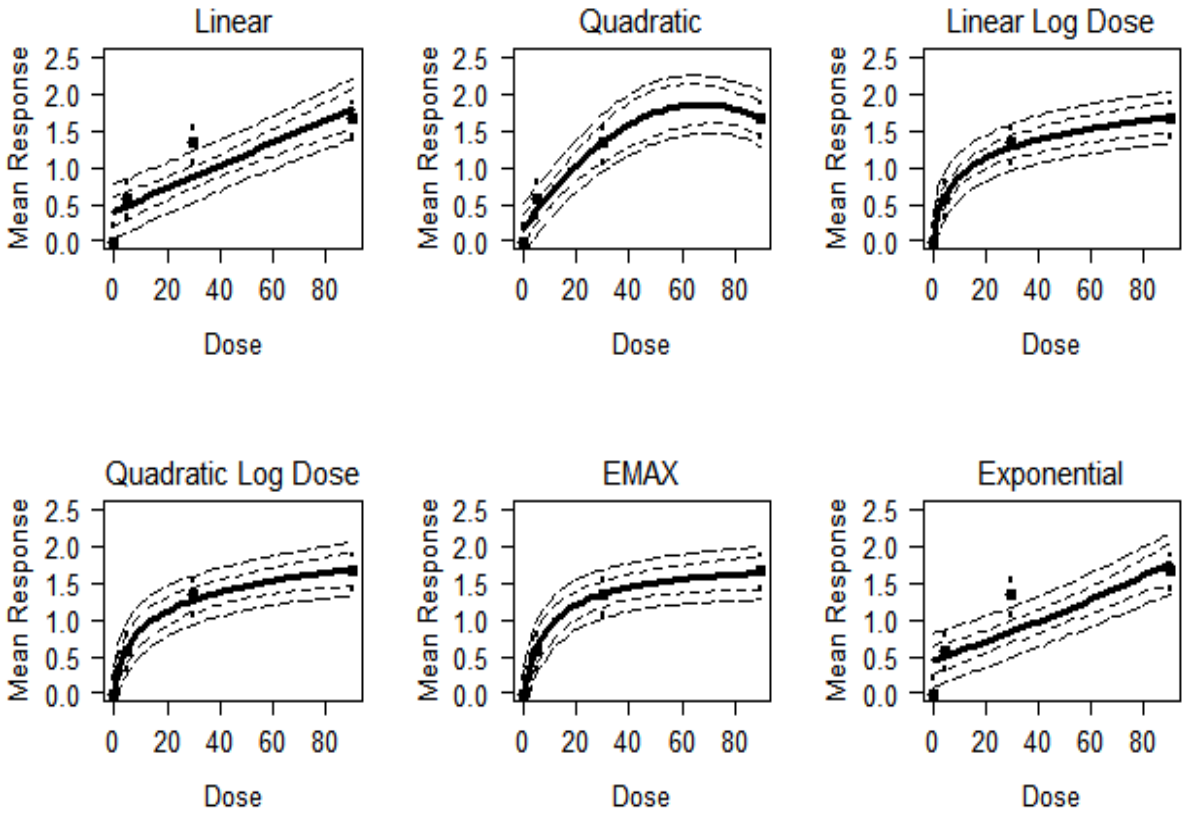




- More details later
- Any dose-response modeling approach should pick this up quickly

Another Example: Correct model not so obvious

- First-in-man study of immunogenicity and safety of a vaccine in healthy adults
- Response = fold increase in immune response over baseline 14 days after administration
- More later



- Linear Log Dose, Quadratic Log Dose, & EMAX provide good fits -- which to choose?
- **Why choose?**

Method

Strategy

- Step 1: Identify candidate D-R models
- Step 2:
  - Fit each model to observed data using conventional MCMC calculations (e.g. STAN, JAGS)
  - Get realizations from posterior distribution of parameters for each model
- Step 3:
  - Get posterior distributions of functions of parameters using part 1 results

Discrete or continuous observed outcomes

Use any probability model

Expected outcome for a dose is a function relating response to dose and, possibly, other covariates

Monotonicity not required

Expected & predicted responses at any set of doses

Doses corresponding to a specified target response

Candidate Dose-Response Models

- Work with weighted linear combination of models commonly used in literature:

1. Linear $f(d; \mathbf{b}) = b_1 + b_2\, d$
2. Quadratic $f(d; \mathbf{b}) = b_1 + b_2\, d + b_3\, d^2$
3. Linear log dose $f(d; \mathbf{b}) = b_1 + b_2 \log(d)$
4. Quadratic log dose $f(d; \mathbf{b}) = b_1 + b_2 \log(d) + b_3 \log(d)^2$
5. EMAX [Sigmoid] $f(d, \mathbf{b}) = b_1 + b_2\, d^{b_4}/(b_3 + d^{b_4})$
6. Exponential $f(d; \mathbf{b}) = b_1 + b_2(\exp(b_3\, d) - 1)$

- Many other models can be used, e.g., cubic splines, fractional polynomials, nonparametric & semi-parametric models, etc.
- Using a set of models with few parameters $\Rightarrow$ wt'd average curve has adequate flexibility, can include trials with few doses
- Not clear that including more than 6 models really is necessary, or that fewer than 6 models would be enough

Posterior Distribution of Expected Response

- Example: Linear Log Dose $f(d;b) = b_1 + b_2 \log(d)$
- MCMC calculations provide an array of realizations from the posterior dist'n of (b_1, b_2)

$$\begin{bmatrix} b_{11} & b_{21} \\ \vdots & \vdots \\ b_{1N} & b_{2N} \end{bmatrix}$$

- Use this array to produce an array of realizations from posterior dist'n of expected responses to a series of doses $d_1, \ldots, d_K$

$$\begin{bmatrix} b_{11} + b_{21}d_1 & \cdots & b_{11} + b_{21}d_K \\ \vdots & \vdots & \vdots \\ b_{1N} + b_{2N}d_1 & \cdots & b_{1N} + b_{2N}d_K \end{bmatrix}$$

- Use the expected response array to get summaries of the posterior distribution of the expected response curve, e.g., mean, credible intervals
- Take a weighted linear combination of the arrays corresponding to the candidate models to get the BMA weighted model

Something Different: First Derivatives

- Same Example: Linear Log Dose $f(d;b) = b_1 + b_2 \log(d)$
- Look at 1st derivative: $f'(d;b) = b_2/d$
- As before, get array of realizations from the posterior dist'n of 1st derivative values at $d_1, \ldots, d_K$

$$\begin{bmatrix} b_{21}/d_1 & \cdots & b_{21}/d_K \\ \vdots & \vdots & \vdots \\ b_{2N}/d_1 & \cdots & b_{2N}/d_K \end{bmatrix}$$

- Use this array to get summaries of the posterior distribution of the first derivative of the expected response curve
- Since sum of derivatives = derivative of sum, weighted sum of model-specific 1st derivative curves = 1st derivative curve of weighted (BMA) model
- Easy PoC evaluation: lower CI of 1st derivative curve > 0
  - Also easy to see where dose-response flattens

Determining Prior Model Weights

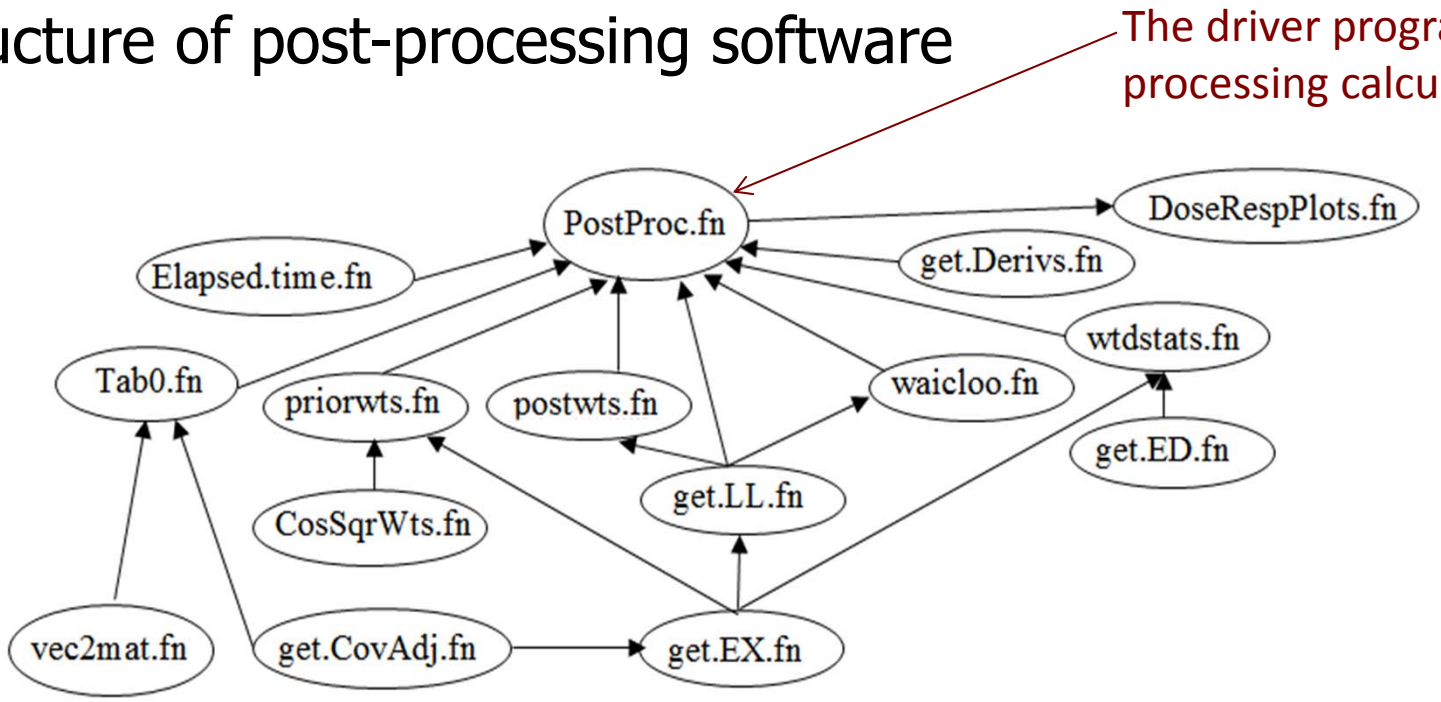
- Clinicians provide explicit weights directly (MCPMod)
- Clinicians provide pairwise ratings of relative preferences for the models (e.g., linear vs EMAX)
  - Construct pairwise preference matrix
  - Weights are elts of eigenvector corresponding to maximum positive eigenvalue. AHP approach uses this method.
- Literature generally assumes uniform weights for models, but this may give undue weights to essentially equivalent models
- Better: Initially assume equal weights for all models, then adjust weights to reflect correlations among predictions provided by each model (Garthwaite *et al*, ANZJS2010, Bka2012)
- This is the default approach, but explicit specification of weights is an option

Identifying Appropriate Doses

- Two ways to proceed (addressing different questions)
- Lower (or upper) quantiles of posterior distribution of responses given doses
  - "What is the least dose that provides 100γ% posterior probability of a response R > r?"
- Posterior distribution of doses given response
  - "Given an observed response R = r, what is the posterior modal dose, and what is the corresponding posterior ci for the dose?"

Software

- Calculations proceed in 3 steps
  - Using MCMC to obtain realizations from posterior distributions of model parameters
  - Post-processing the MCMC output to produce summary datasets and tabulations
  - Graphical displays that drive conclusions
- Structure of post-processing software



Examples

Example 1: Sigmoid Dose-Response

Sigmoid DR Curve (Ruberg JBS1995)

- Data from article (digitized)
- Prior weights

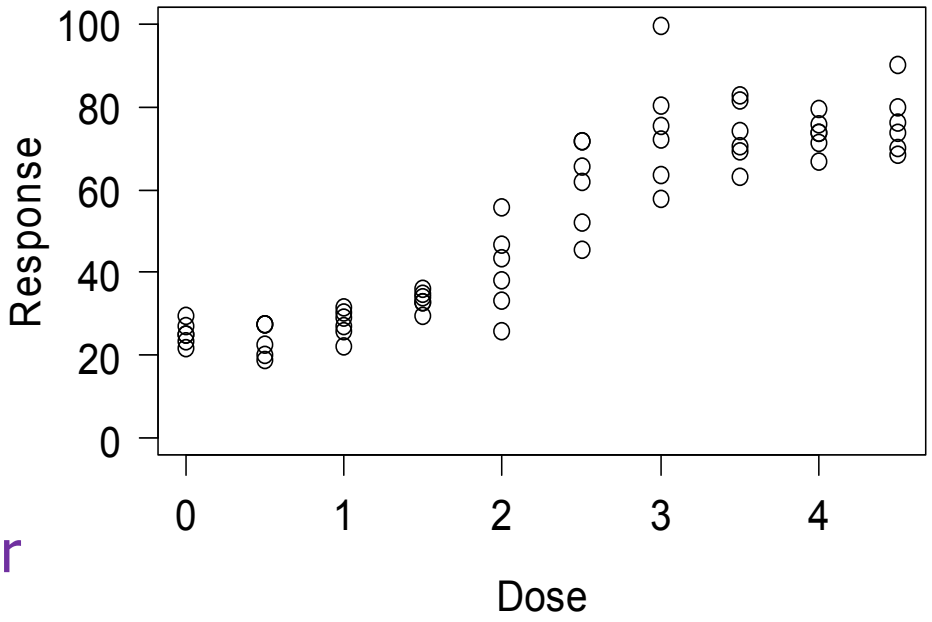
0.02, 0.02, 0.34, 0.05, 0.35, 0.22

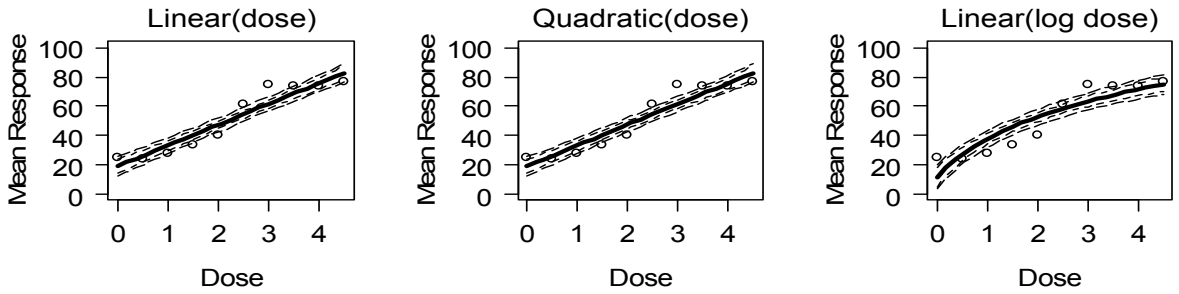
- Posterior weights

0, 0, 0, 0, 1, 0

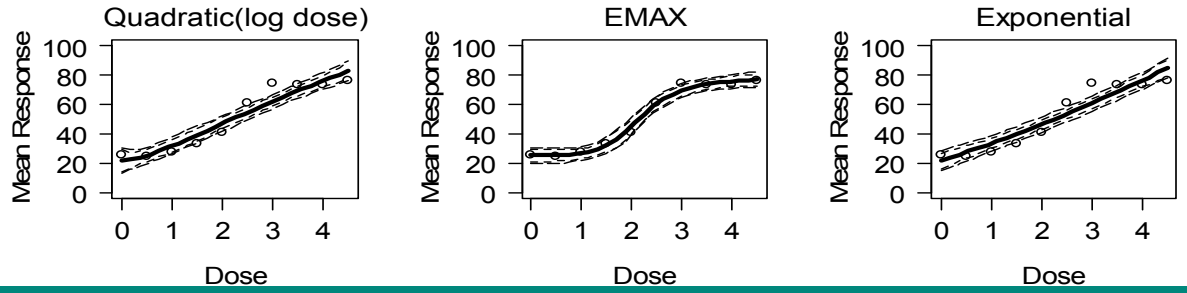
normal or t

Model likelihoods → 'correct' post weights regardless of prior





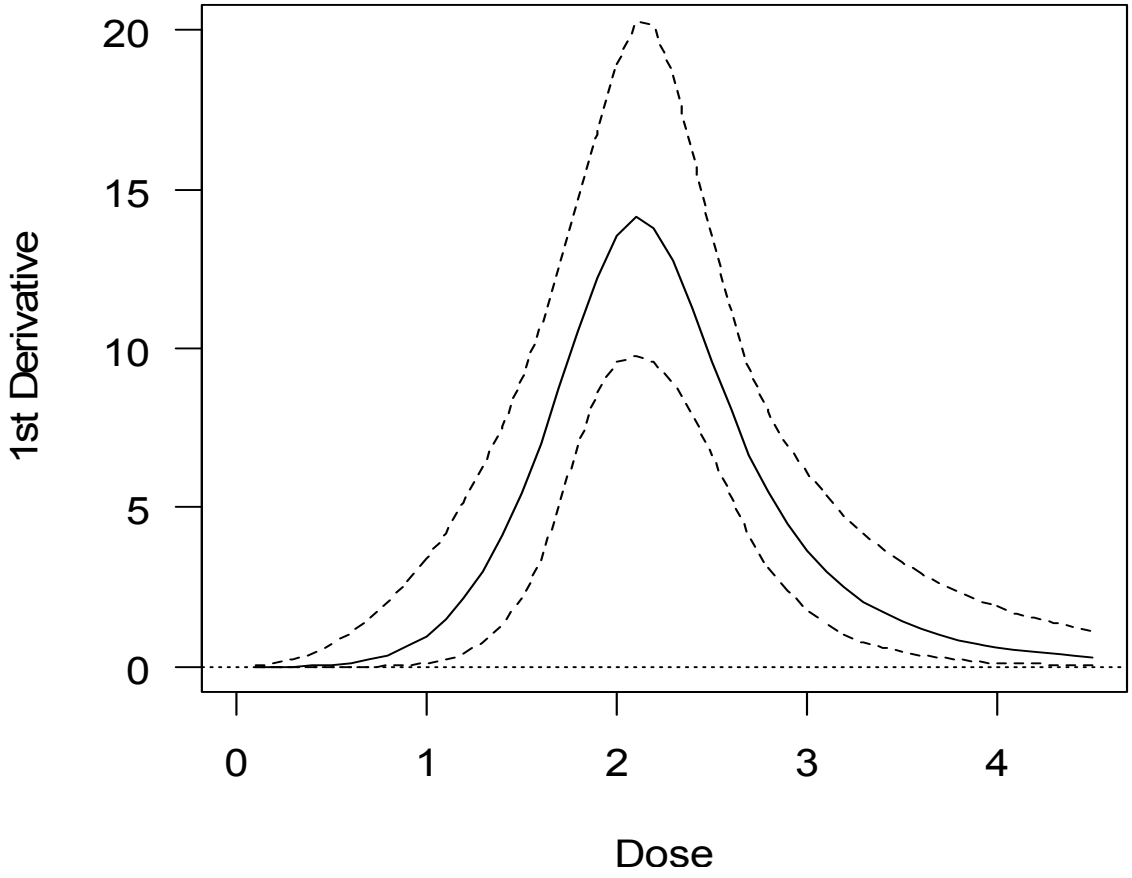
Expected posterior DR trajectories & 95% CI



EMAX (sigmoid) model provides best fit to data especially with t(3) likeihood

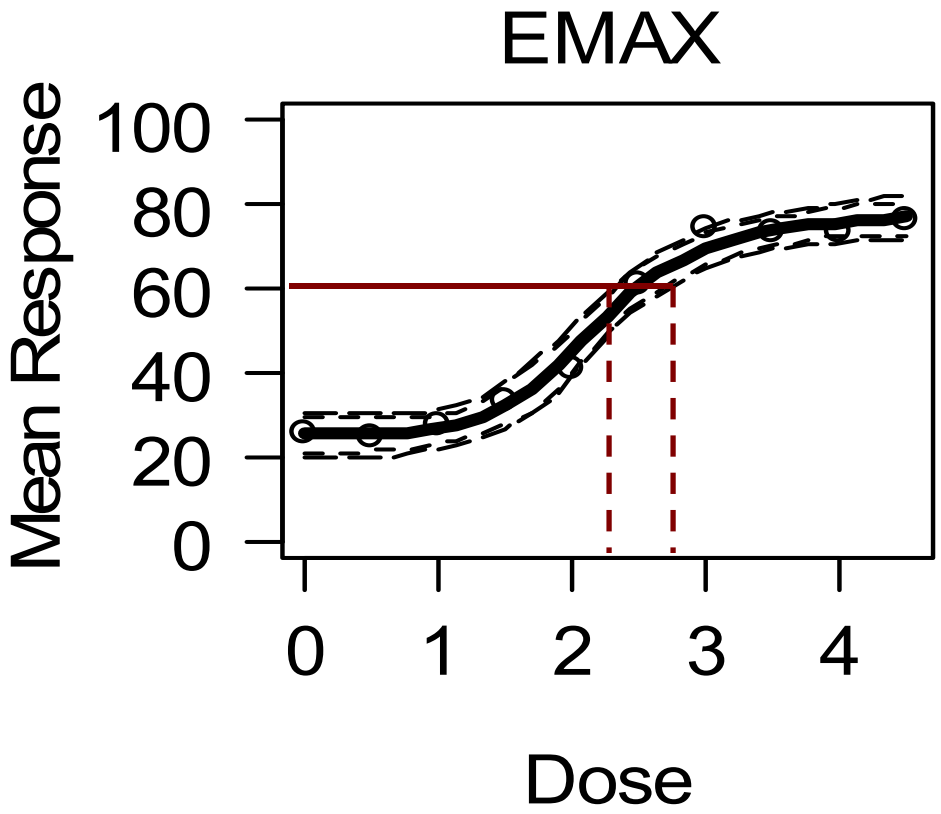
First Derivative of the Optimal Weighted D-R Function

- Posterior distribution of 1st derivatives of weighted dose response function with 95% CI bounds
  - Lower 95% bound > 0 ⇒ positive dose-resp relationship



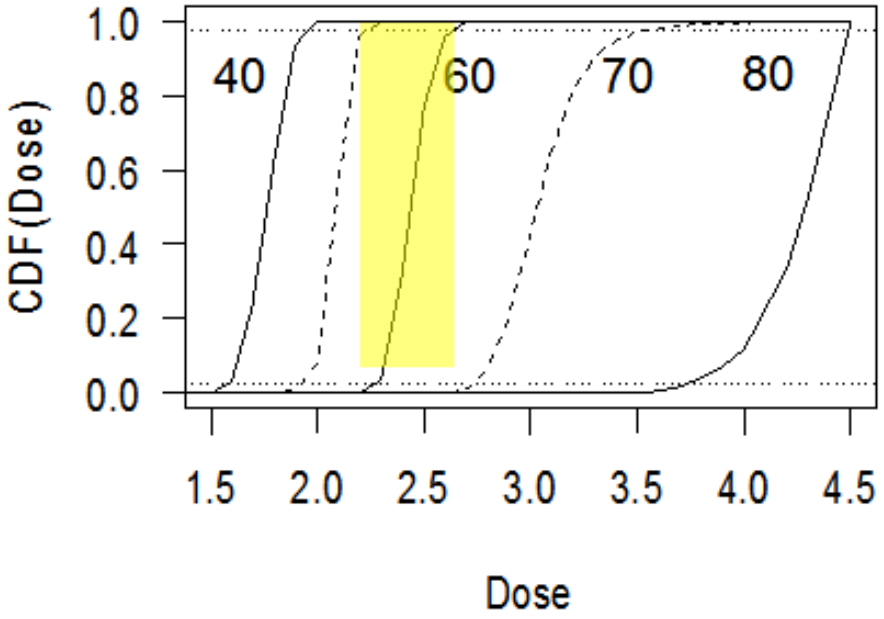
Target Dose (1): Least dose ∋ $P_{post}(R > r \mid dose) \geq \gamma$

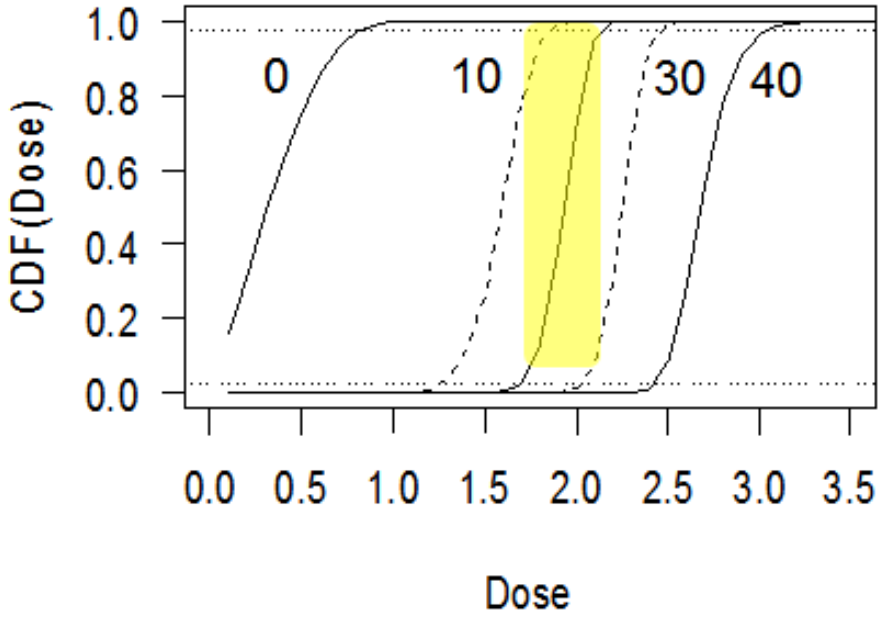
- Quantiles of posterior dist'n of responses given doses



Target Dose (2): CDF_{post}(Dose | Response)

- Posterior cdfs of dose as a function of outcome target, either as actual expected response or as expected difference from 0 dose





- E(Response) = 60 $\Rightarrow$ 95% post CI for dose = (2.25, 2.75)
  - Lower doses are unlikely to give same expected response
  - Higher doses may present an elevated AE risk
- Same calc for AE risk $\rightarrow$ dose choice balances benefit & risk

Example 2: Binary Outcomes

Summary of Data Source

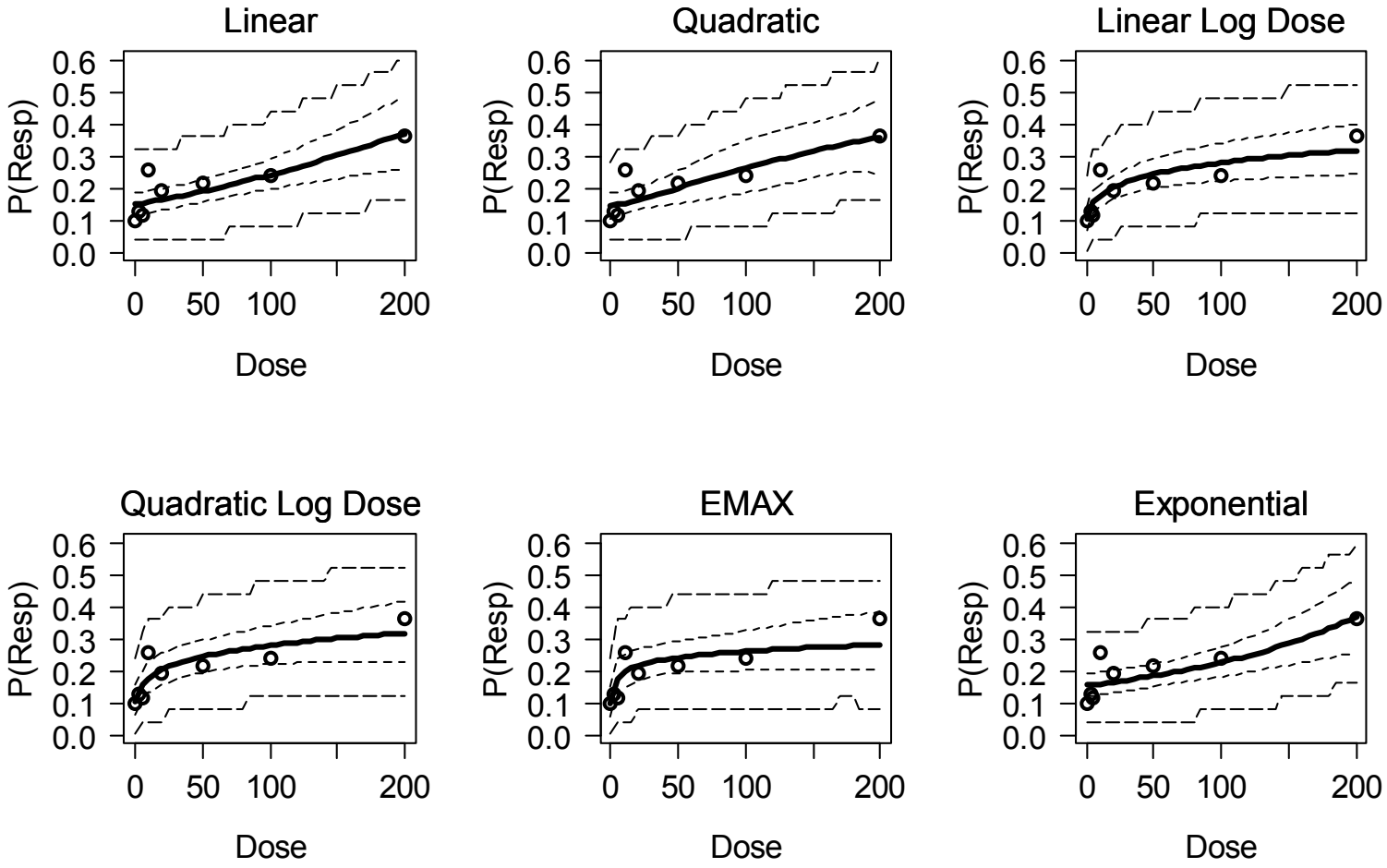
- Response = "pain free 2 hours post dose" from a randomized placebo-controlled trial of a compound for treating acute migraine; 7 active doses [1.usa.gov/28Xd9Hr]
- Key question: Which (if any) dose to carry forward?
- Data summary (Diff & OR are comparisons with 0 dose)

Dose	n	X	$\mathbf{p_{Obs}}$	E(p)	$p_{.025}$	$p_{.975}$	$\mathbf{Diff_{Obs}}$	E(Diff)	$Diff_{.025}$	$Diff_{975}$	$\mathbf{OR_{obs}}$	E(OR)	$OR_{.025}$	$OR_{.975}$
0	133	13	**0.10**	0.10	0.06	0.16	**NA**	NA	NA	NA	**NA**	NA	NA	NA
2.5	32	4	**0.12**	0.13	0.04	0.27	**0.03**	0.04	-0.07	0.18	**1.3**	1.3	0.4	4.1
5	44	5	**0.11**	0.11	0.04	0.23	**0.02**	0.02	-0.08	0.14	**1.2**	1.2	0.4	3.3
10	63	16	**0.25**	0.25	0.16	0.37	**0.16**	0.16	0.05	0.28	**3.1**	3.2	1.4	7.1
20	63	12	**0.19**	0.19	0.11	0.3	**0.09**	0.10	-0.01	0.21	**2.2**	2.2	0.9	5.1
50	65	14	**0.22**	0.22	0.13	0.32	**0.12**	0.12	0.01	0.23	**2.5**	2.6	1.1	5.7
100	59	14	**0.24**	0.24	0.14	0.36	**0.14**	0.14	0.03	0.26	**2.9**	2.9	1.3	6.6
200	58	21	**0.36**	0.36	0.25	0.49	**0.26**	0.26	0.14	0.40	**5.2**	5.2	2.4	11.6

Some Posterior Results

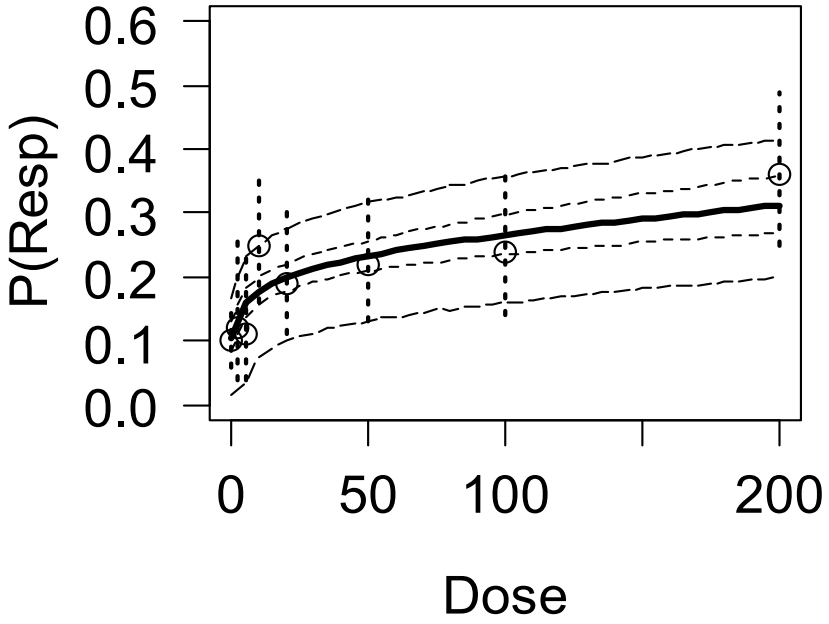
- Posterior model weights:

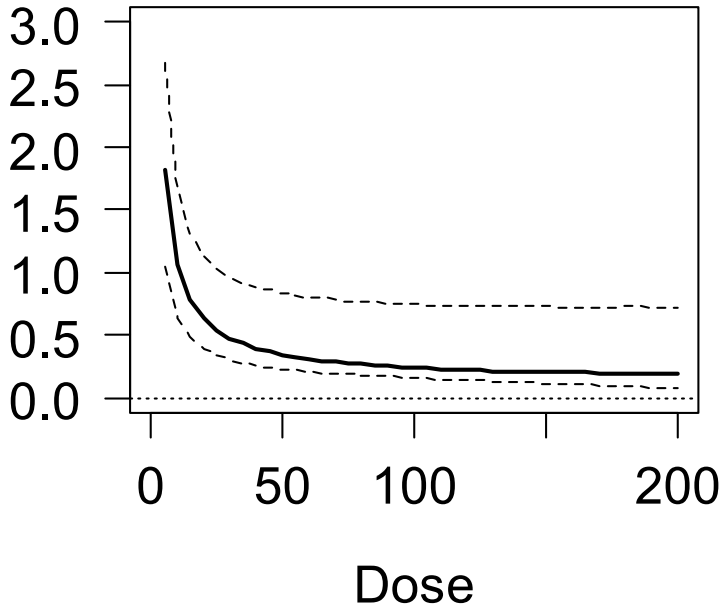
Linear	Quadratic	Linear Log	Quadratic Log	EMAX	Exponential
0.04	0.06	0.36	0.17	0.30	0.07



BMA Weighted Model Summary

- Posterior expected dose-response fn and 1st derivative ($\times$ 0.01)

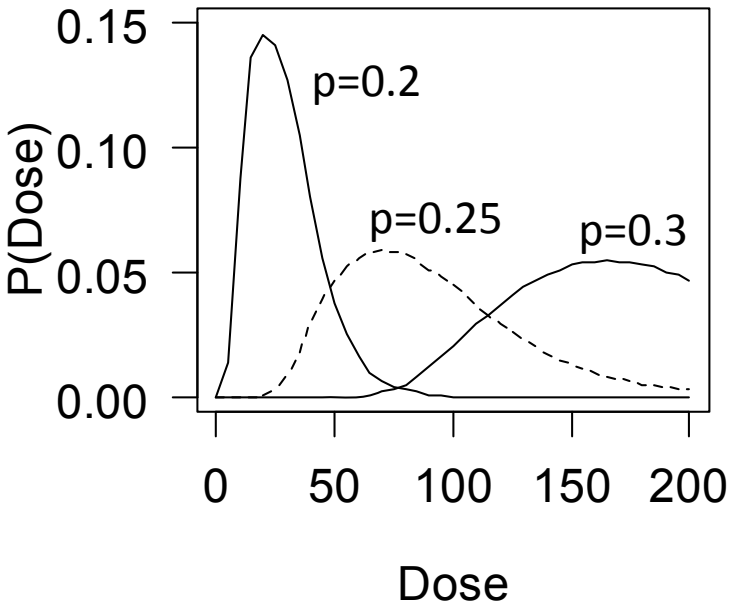


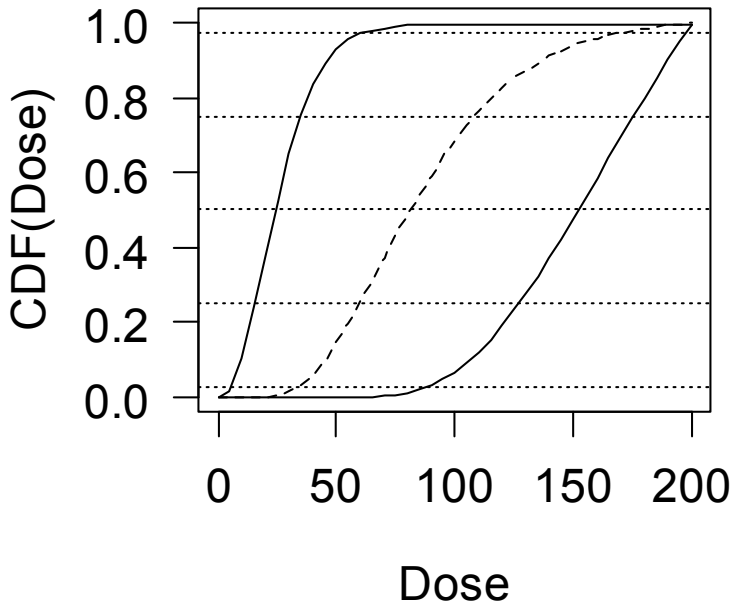


- There is a monotonic dose-response relationship
- Not much additional effect past doses of 50 or 100

Doses to Consider Carrying Forward

- Posterior distributions of doses as a function of target response



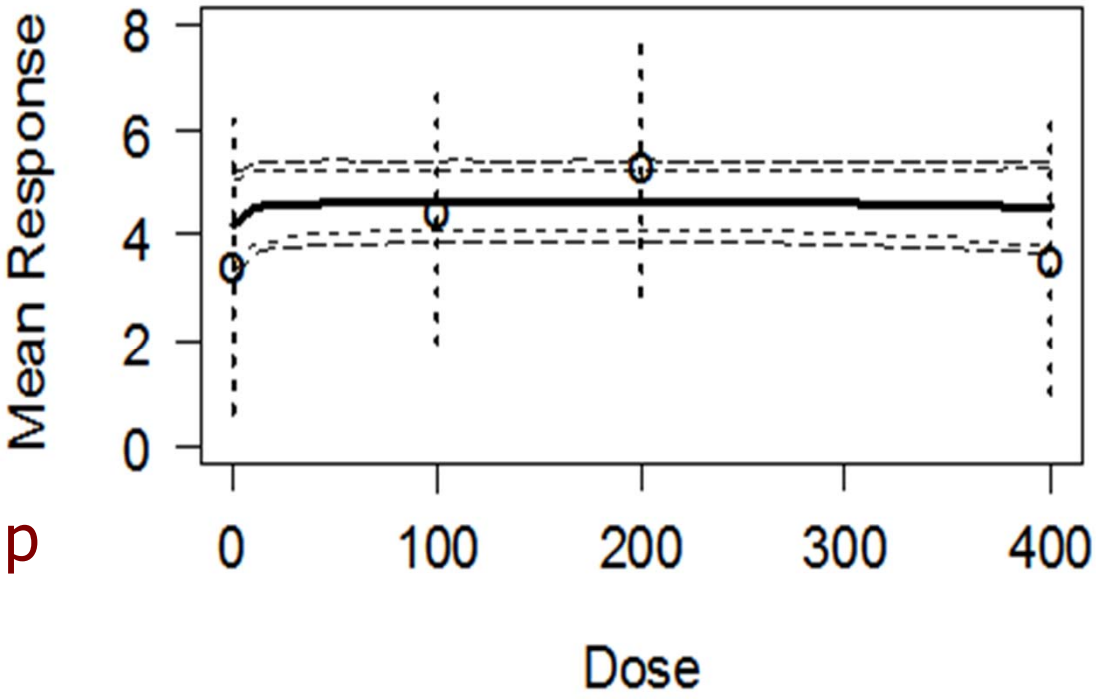


- Target response rate = 0.25 $\Rightarrow$ 95% CI for corresponding dose is (40,150), and interquartile range is (55,100)
- Straightforward to carry out calculation if target response is expressed as difference from zero dose or odds ratio relative to zero dose

Example 3: DR Models with Covariates

Continuous Response, with Covariates

- Response = % chng from bsln in FEV1 after 8 wks of trt in a trial evaluating asthma treatment
- Posterior means & 95% CI for predicted mean response by dose, weighted dose-response curve
- Not much evidence of a dose-response relationship



- Prior weights:

  0.19, 0.19, 0.15, 0.15, 0.15, 0.17

- Posterior weights:

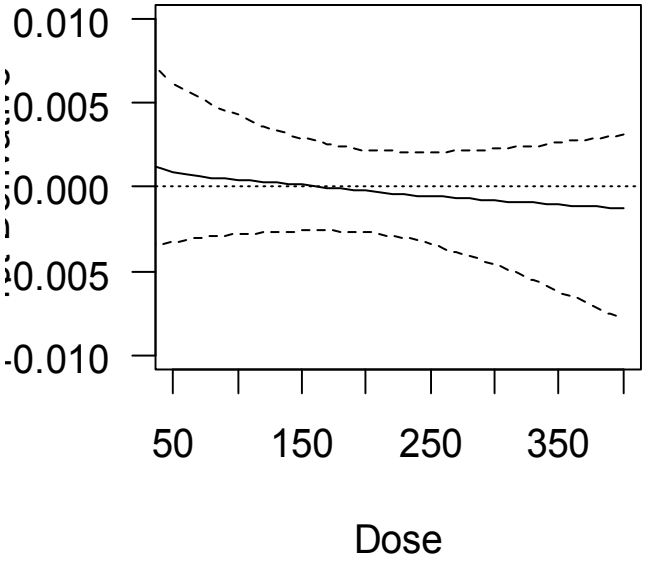
  0.14, 0.18, 0.16, 0.17, 0.19, 0.17

Covariates:

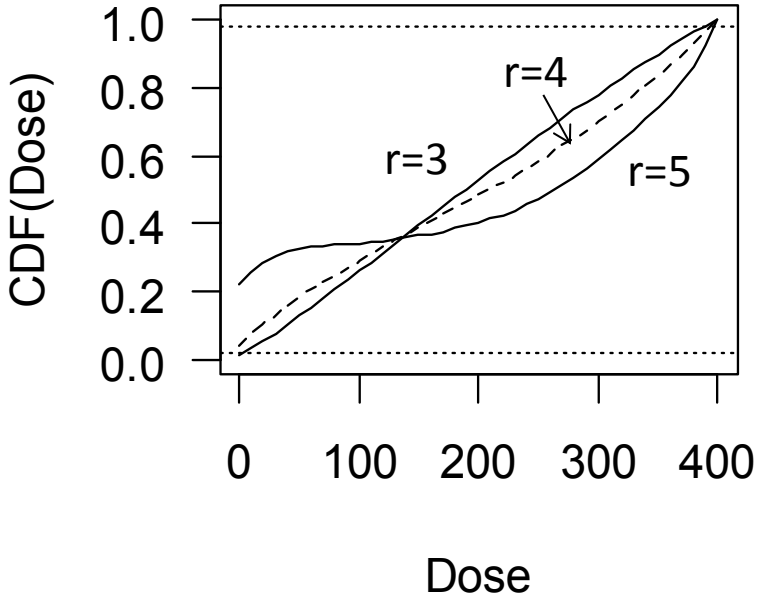
Age Category (<,≥ 7)	(-2.8, 4.1)
Bsln FEV1	(-8.5, -3.2)

Derivatives and Doses

- Posterior mean and 95% CI bounds for the 1st derivative of the weighed dose-response curve
  - Not even a monotonic d-r relationship



- Prior probabilities of doses based on expected response



- No dose selection guidance

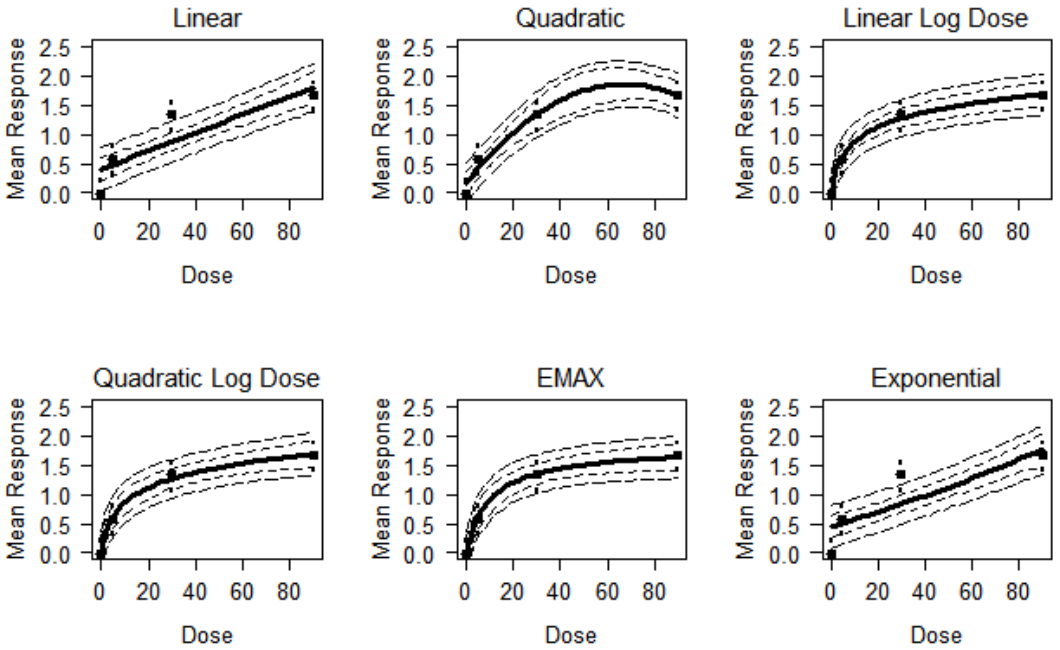
Example 4: Multiple ‘Adequate’ Models

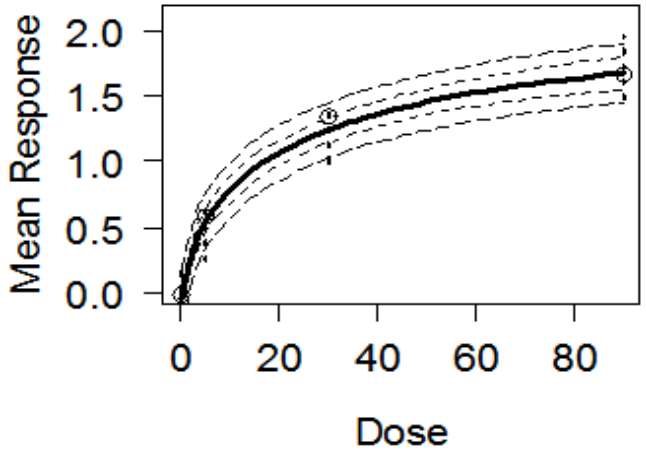
Multiple Satisfactory Models

- First-in-man study of immunogenicity and safety of a vaccine in healthy adults
- Response = fold increase in immune response over baseline 14 days after administration
- Prior weights

  0.17, 0.16, 0.16, 0.15, 0.19, 0.19
- Posterior weights

  0, 0.06, 0.33, 0.27, 0.34, 0
- Linear log dose, Quadratic log dose, EMAX provide best fits
- Optimal weighted Bayes DR curve is best

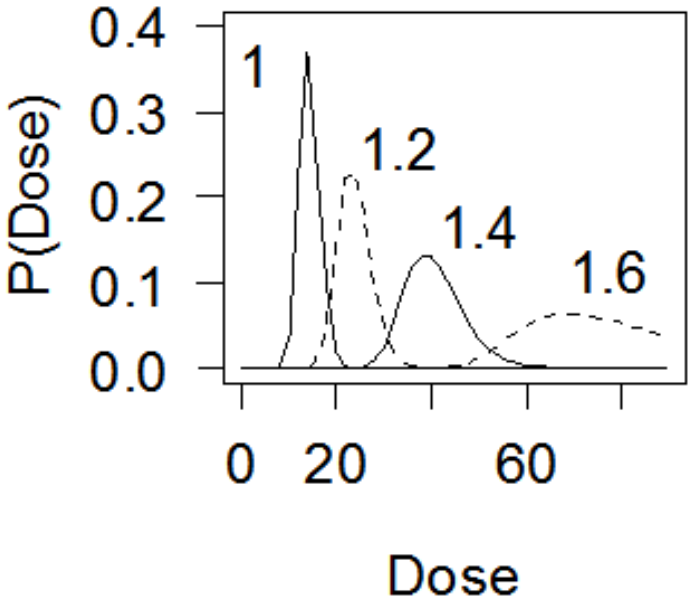




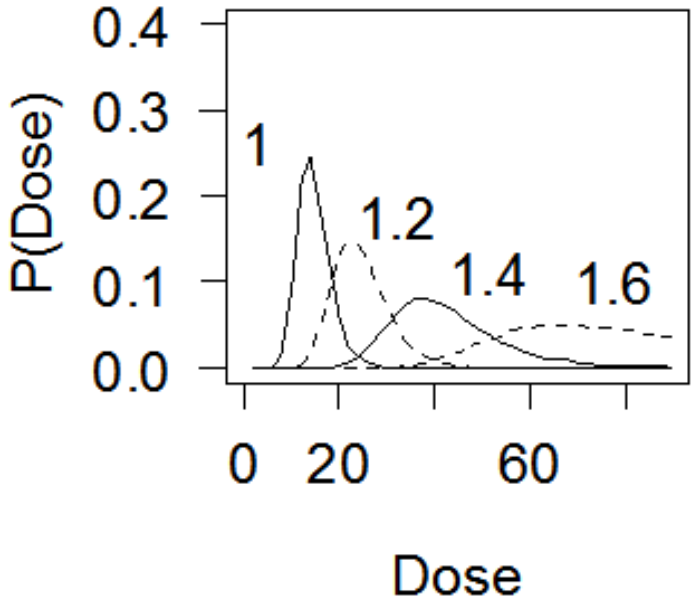
Target Dose Selection

- Posterior probabilities of doses, based on

Expected response (r)



Difference from 0 dose



- Objective: 1.4 fold difference from zero dose response
  - Modal dose ~ 40 μg
  - 95% credible interval ~ (22, 72)
  - Incorporates posterior variability of responses to zero dose

Example 5: Binary Responses with Random Center Effects

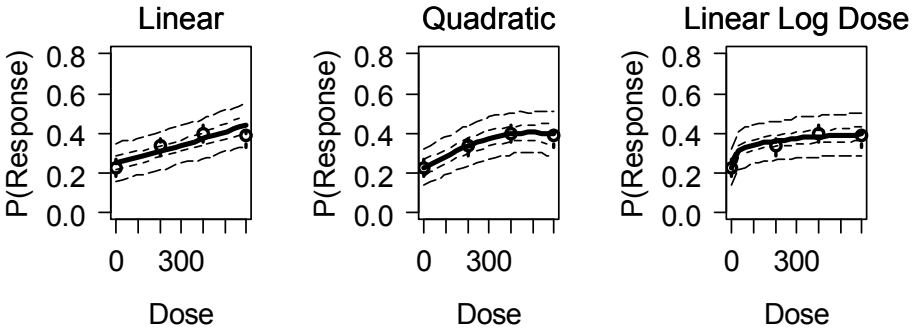
Ohlssen & Racine (JBS2015)

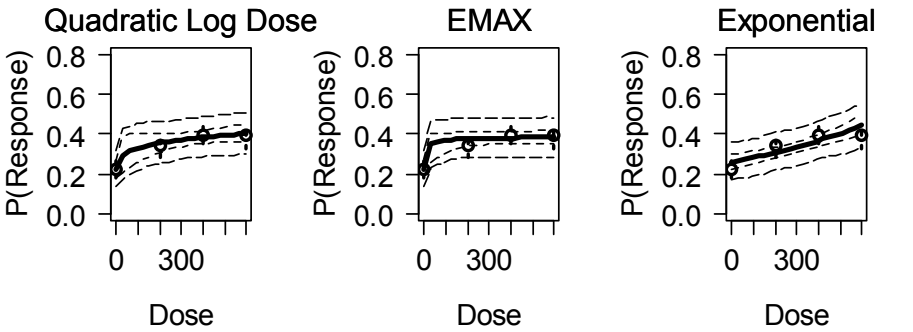
- Trial in 3 centers comparing vimpatin vs placebo as adjunct Tx for treating partial-onset epileptic seizures in patients > 15 yrs
- Response = $\geq$ 50% reduction in seizure frequency from bsln post 12 wks of maintenance after 4-6 wks forced titration
- O&R used nonparametric monotone regression for analysis
- Summary of data
  - Wider bounds for response prob with random center effect
  - Less pronounced for difference or odds ratio

		Observed		No Center Effect Posterior			Random Center Effect Posterior		
Dose	n	X	p_{Obs}	p_{Exp}	LB	UB	p_{Exp}	LB	UB
0	360	81	0.23	0.23	0.18	0.27	0.23	0.16	0.35
200	267	91	0.34	0.34	0.29	0.40	0.34	0.25	0.49
400	470	186	0.4	0.4	0.35	0.44	0.41	0.31	0.55
600	203	80	0.39	0.39	0.33	0.46	0.41	0.31	0.57

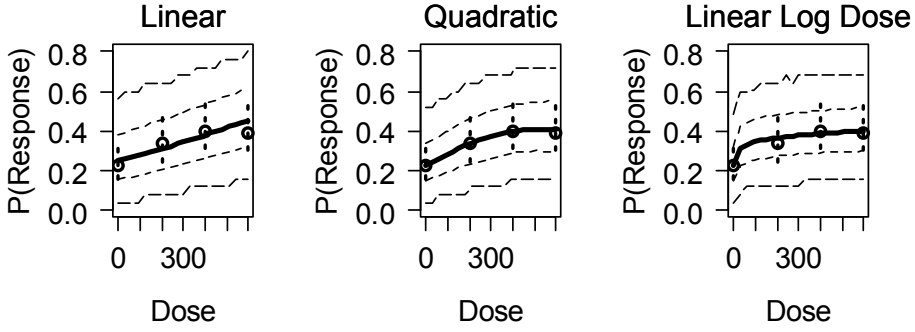
Influence of Random Center Effect

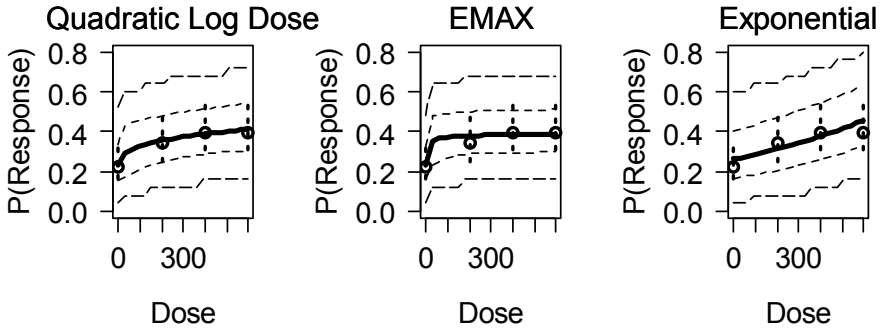
No Center Effect

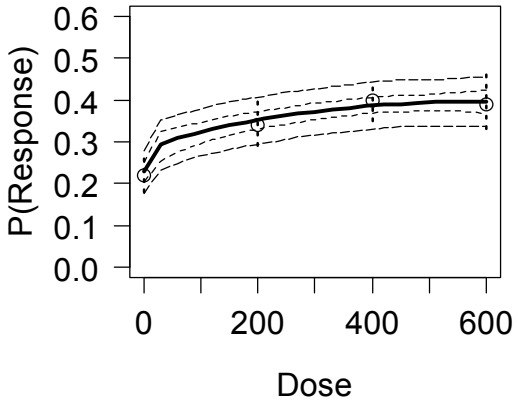




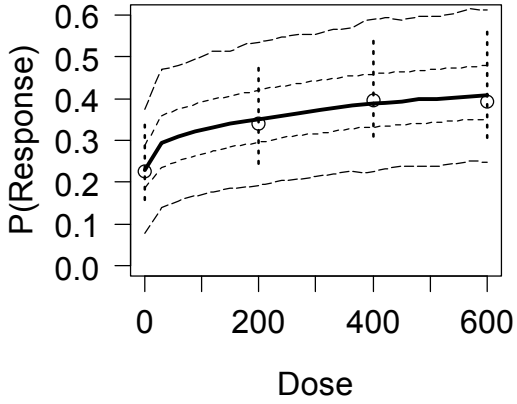
Random Center Effect







Averaged Models



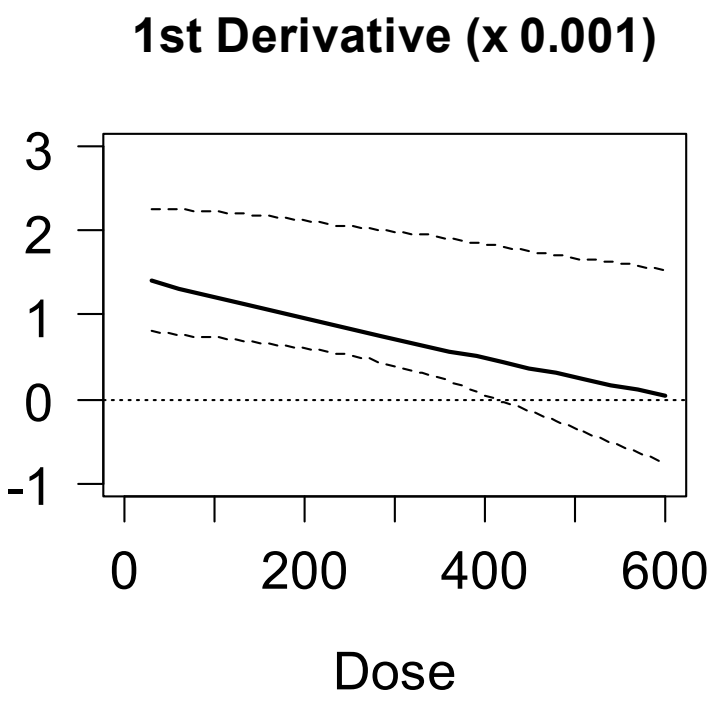
Posterior Model Weights

Prior Weights	Linear	Quadratic	Linear Log	Quadratic Log	EMAX	Exp
Garthwaite	0.09	0.25	0.04	0.22	0.35	0.05
Uniform	0.06	0.21	0.26	0.22	0.21	0.04
'Pessimistic'	0.15	0.18	0.22	0.18	0.17	0.11

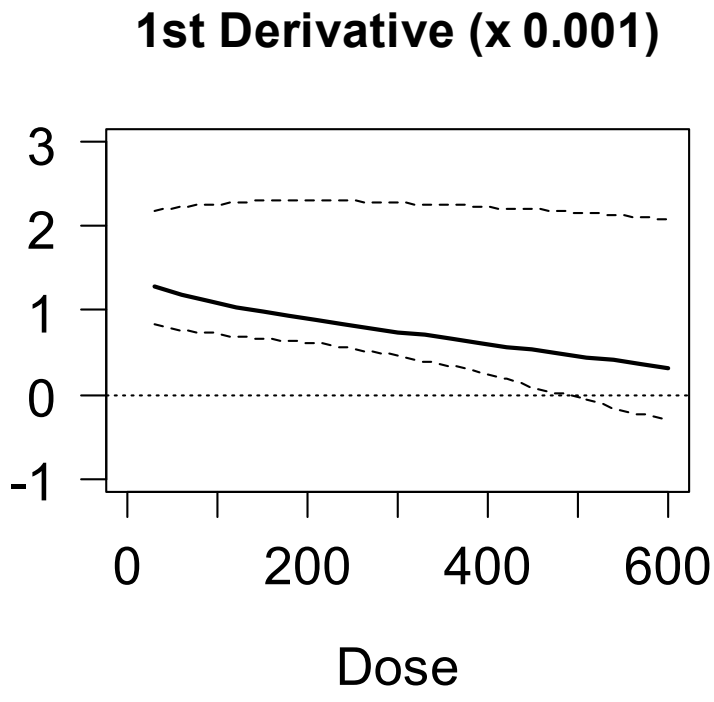
- Posterior weights not very sensitive to prior weights, weighted log likelihood essentially the same for all prior weight choices

First Derivatives – Is there a D-R Relationship?

No Center Effect



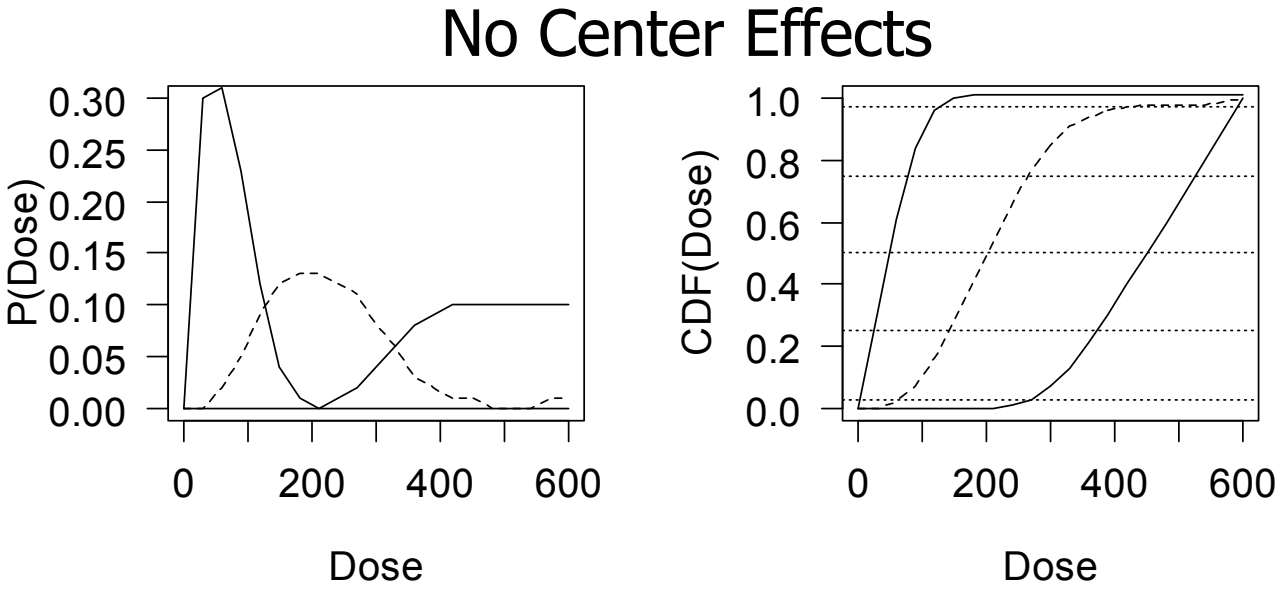
Random Center Effect

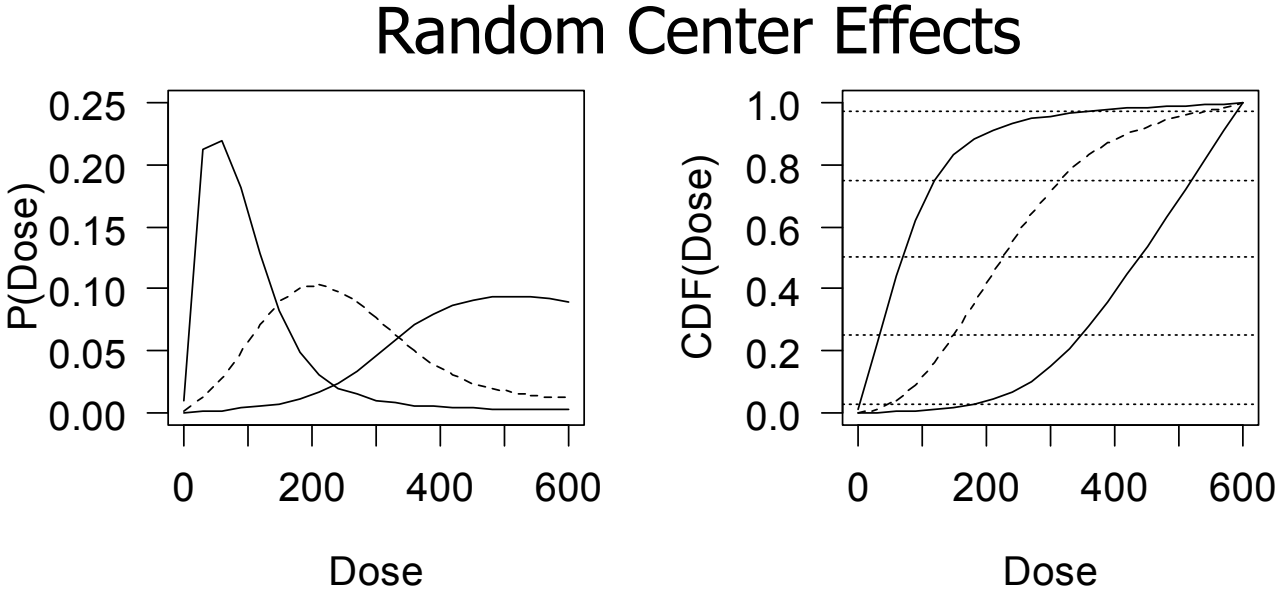


- Response may increase with dose up to about 400 mg, but appears to flatten out thereafter

What Dose to Pursue?

- Target response rates are (L to R) 0.3, 0.35, 0.4
- Ignoring center effects ⇒ 200 mg likely to produce response rate > 30%
- Including center effects ⇒ chance of 30% response rate with 200 mg is ~ 50%





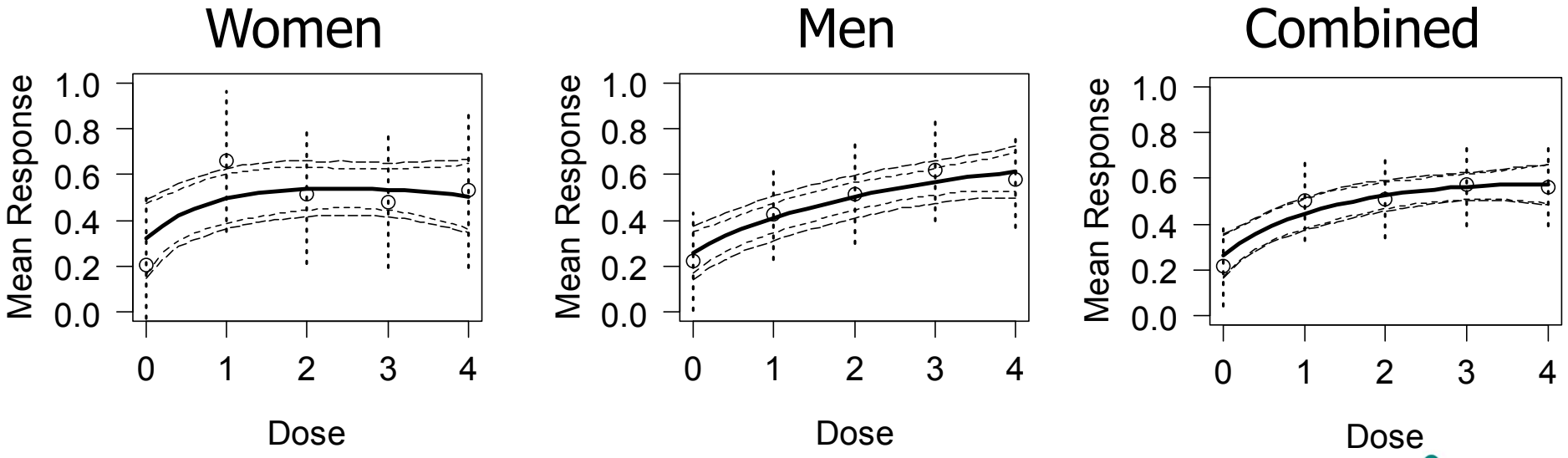
Example 6: Covariate Adjustments to All Parameters

'IBScovars' data set from MCPmod R package

- Dose-ranging trial of 4 doses (plus pbo) of a compound for treating irritable bowel syndrome
- Posterior model weights:

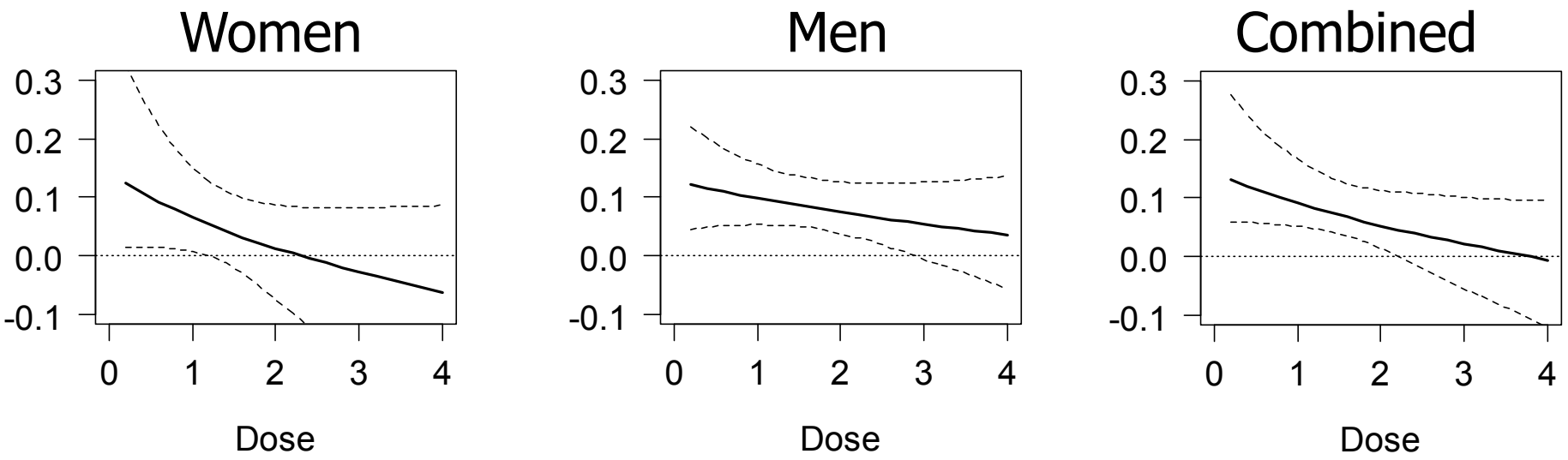
	N	Linear	Quadratic	Linear Log	Quadratic Log	EMAX	Exponential
Women	118	0.16	0.18	0.02	0.42	0.04	0.18
Men	251	0.18	0.28	0.01	0.27	0.06	0.20
Combined	369	0.13	0.25	0.00	0.45	0.06	0.11

- Weighted dose-response curves



Is There a Monotone Dose – Response Relationship?

- First Derivatives



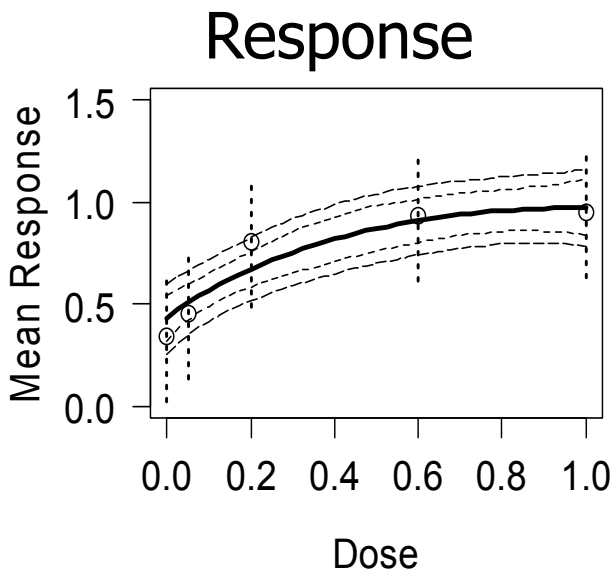
- Response for women unlikely to increase with dose after Dose 2, but response for men appears to be real, possibly to Dose 4
- Log likelihoods for weighted models not sensitive to inclusion/exclusion of gender effect or choice of prior model weights

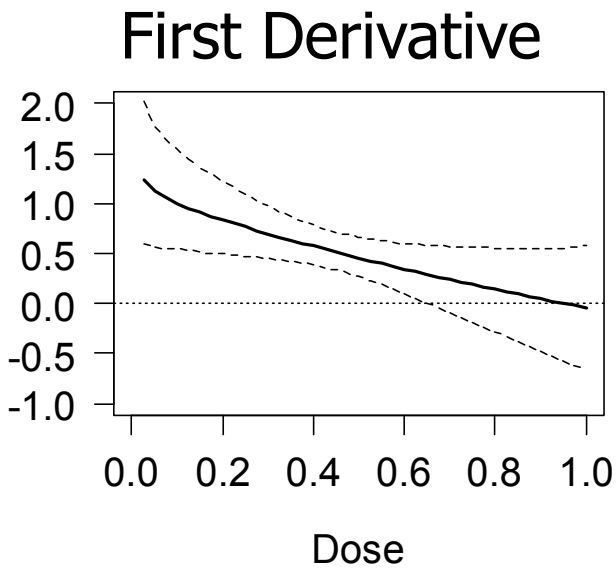
Example 7: Dose Range Limitation

Data and Summary

- 'biom' dataset in MCPmd R packagePosterior model weights
- Posterior weighted dose-response curve and 1st derivatives

Linear	Quadratic	Linear Log	Quadratic Log	EMAX	Exponential
0.08	0.24	0.13	0.30	0.19	0.06

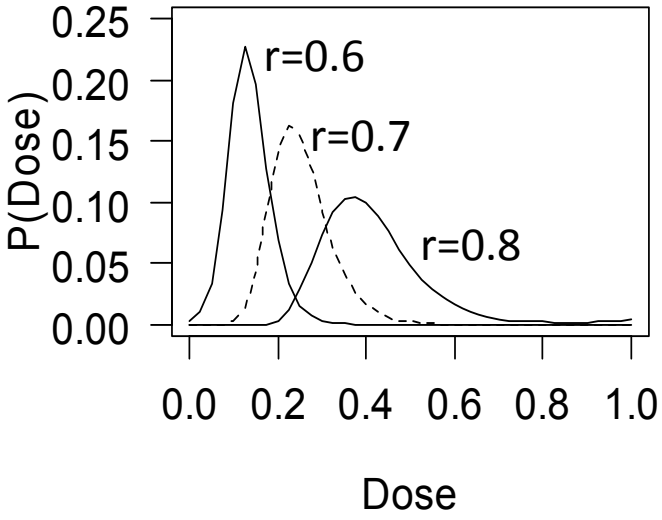


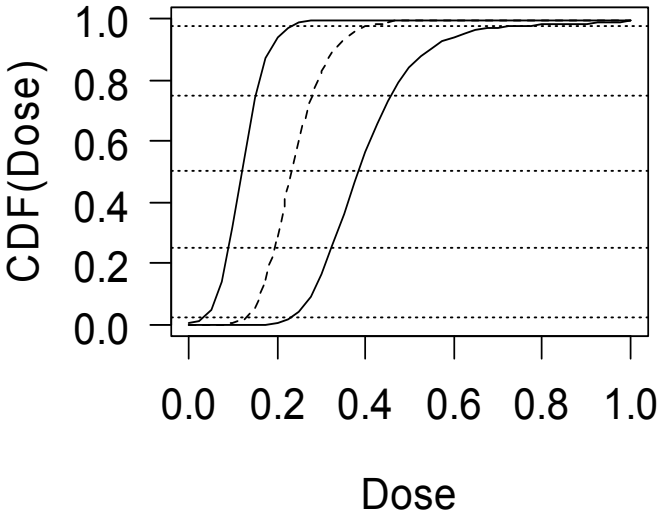


- Gutjahr & Bornkamp [Bcs2017] demonstrated signficant d-r relationships with linear, exponential, and EMAX models
- Probably is supportable only for doses < 0.6 given 1st deriv.

Which Doses to Pursue?

- Target response of 0.6 or 0.7 seems to provide reasonable dose selection guidance





Example 8: Slopes from a Random Effects Linear Model

Data and Summary

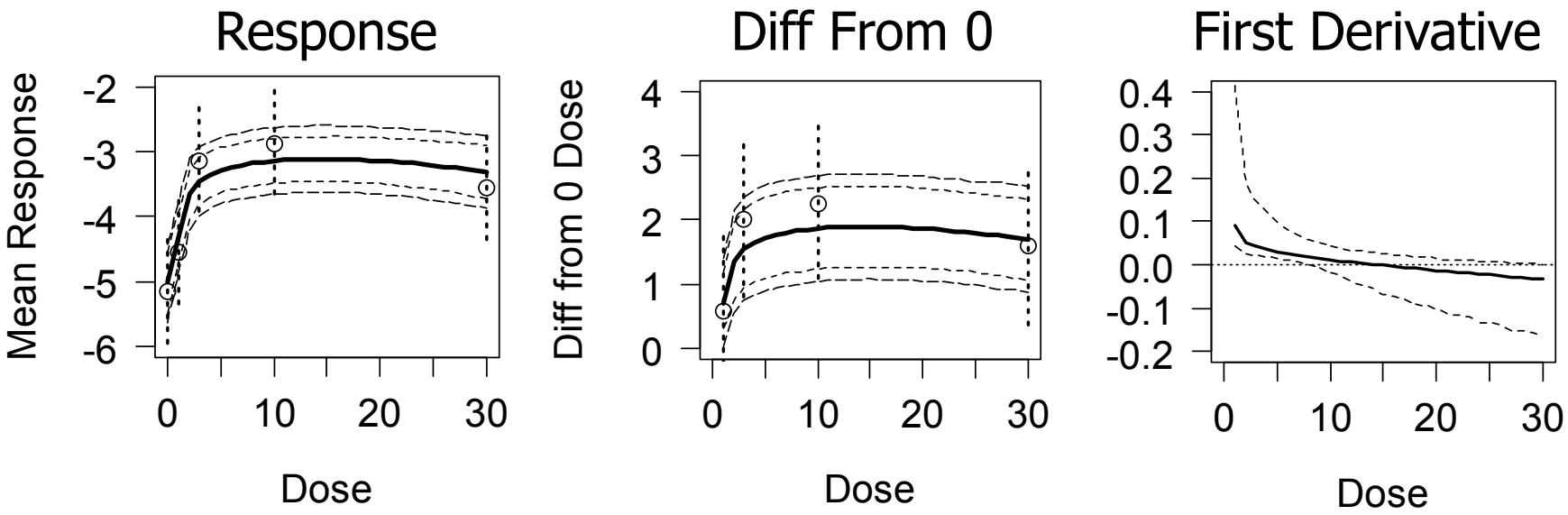
- Same as Example 4.1 in Pinheiro *et al* [StatMed2014]
- Evaluate effect of various doses of a new drug on disease rate progression ≡ slope of linear regn fitted to a patient's obsns
- 100 patients, 5 functional scale measurement times, treated with 0, 1, 3, 10, or 30 mg
- Step 1: Fit mixed effects linear regns to patients' sequences, easy to do using lme function in nlme R package to get patient-specific slopes
- Posterior weights for the various models

Linear	Quadratic	Linear Log	Quadratic Log	EMAX	Exponential
0.00	0.10	0.02	0.39	0.49	0.00

- Model fit evaluations confirm that quadratic log dose and EMAX provide the best fits

Posterior Summary

- Posterior dose-response function and 1st derivatives



- Pinheiro *et al* estimate dose needed to get 1.4 unit improvement relative to pbo as 2.13 or 2.15

- Better choice might be 5 or 6 units because of variability, to guard against inadequate dose response

Comments and Discussion

The Message

- Bayesian Model Averaging is an efficient and flexible tool for
  - Characterizing dose-response relationships
  - Finding 'correct' doses to carry forward
- Key Attributes
  - No need to identify a 'best' model
  - Multiplicity adjustments not needed
  - Flexible distributional structure for data likelihood
  - Direct assessment of PoC using small samples
  - Identification of dose range corresponding to specified response targets
- Definitive graphical displays of analysis results simplifies communication
- Software (reasonably user friendly) available for calculations

BMA Analyses Useful for:

- Inferences about the response at each dose level
- Separate objectives: (a) Determining if $\exists$ a D-R relationship
  (b) Identifying appropriate dose ranges
  - Lower quantile of posterior dist'n of 1st derivative > 0 $\Rightarrow$ reasonable to conclude a D-R relationship exists
  - Establish or rule out PoC with modest trials
  - Use predictive distributions of future responses to inform definitive trials of specific doses
- No need for approximations or assumptions of asymptotic behavior
- Include as many models as are clinically sensible
- Posterior distributions of doses corresponding to target outcomes can be used to determine doses to carry forward

Further Comments

- Just identifying doses that differ significantly from control may not be best dose finding strategy
- Better: quantify the anticipated responses to a set of doses, use results to determine dose range that gives target outcomes
  - High enough to provide target outcome
  - Low enough to minimize toxicity risk
- Determining appropriate dose depends on value of corresponding responses, which depends on
  - Effectiveness
  - Toxicity potential
  - Appeal/convenience of dosage regimen
  - Competition
  - Production cost

Comparison of Bayes & MCPMod (1)

MCPMod	Bayes
Clinical team: decide on the core aspects of the trial design, specify M plausible dose-response models and K doses to include in the trial.	
Determine model-specific 'optimal' dose-response contrasts & sample sizes needed to detect a dose-response relationship	Identify likelihoods & parameters for each model Determine prior model weights (various approaches possible)
Fit each model to observed data using ANOVA or regression, estimate expected response at each dose, assuming asymptotic normality of parameter estimates	Use MCMC to get realizations from joint posterior distns of each model's parameters , obtain realizations of functions of the parameters.
Use 'optimal' contrasts to identify dose-response relationships for each model. Overall test uses max of model-specific test statistics, critical values assume multivariate normality	No testing. Estimates of functions of parameters include credible intervals. Could base positive dose-response conclusion on posterior dist'n of first derivatives of D-R functions.

Comparison of Bayes & MCPMod (2)

MCPMod	Bayes
If significant max z value, identify a 'reference set' of the models with z-statistics large enough to reject the null hypothesis of zero 'optimal' contrast values.	No need to select a 'best' model. Bayesian model averaging → optimum estimate that combines input from all models.
Use inverse regression based on a 'best' model or a weighted average of models to estimate target doses for further development; determine precision using bootstrapping	Determine posterior probabilities of doses corresponding to responses falling in a specified interval. Use with utility functions or other information to guide dose selection.

Post Processing

- N realizations (MCMC) from joint posterior dist'n of parameters for model m (m = 1, …, M)

$\mathbf{b}^{(m)} =$

MCMC Realization	Model m Parameter(s)
1	$\vec{b}_1^{(m)}$
2	$\vec{b}_2^{(m)}$
⋮	⋮
N	$\vec{b}_N^{(m)}$

- Look at posterior distributions of functions of the elements of the $\vec{b}_i^{(m)}$

MCMC	Functions of Model m Parameters
1	$\mathbf{y}_1^{(m)} = f_1(\vec{b}_1^{(m)}),\ f_2(\vec{b}_1^{(m)}), \cdots, f_S(\vec{b}_1^{(m)})$
2	$\mathbf{y}_2^{(m)} = f_1(\vec{b}_2^{(m)}),\ f_2(\vec{b}_2^{(m)}), \cdots, f_S(\vec{b}_2^{(m)})$
⋮	⋮
N	$\mathbf{y}_N^{(m)} = f_1(\vec{b}_N^{(m)}),\ f_2(\vec{b}_N^{(m)}), \cdots, f_S(\vec{b}_N^{(m)})$

- $\mathbf{Y}^{(m)}$ = array with rows $\mathbf{y}_1^{(m)}$, …, $\mathbf{y}_N^{(m)}$ = N realizations from the joint posterior dist'n of functions of the parameters

- Expected responses at various doses → dose distributions
- Likelihood for obs'd outcomes → posterior model weights
- Use weights to calculate Bayes averages of model-specific quantities

Determining Prior Wts using AHP

- Pairwise comparisons of models by relative importance judged by clinicians
  - 1 = equal, 2 = slightly more, 3 = moderately more, 5 = strongly more important

MODEL	Linear	Quadratic	Log Linear	Log Quadratic	EMAX	Exponential
Linear	1	0.33	0.5	0.33	0.2	0.33
Quadratic	3	1	1	1	0.2	0.33
Log Linear	2	1	1	0.33	0.2	0.33
Log Quadratic	3	1	3	1	0.2	0.33
EMAX	5	5	5	5	1	5
Exponential	3	3	3	3	0.2	1

- Model weights: 0.05, 0.09, 0.07, 0.11, 0.48, 0.20
  - Calculate as normalized eigenvector corresponding to maximum positive eigenvalue

Determining 'Effective' Doses (details)

- Matrix $\mathbf{Y}^{(m)}$ of functions of MCMC realizations of model parameters

MCMC	Functions of Model m Parameters
1	$\mathbf{y}_1^{(m)} = f_1(\vec{b}_1^{(m)}),\ f_2(\vec{b}_1^{(m)}), \cdots, f_S(\vec{b}_1^{(m)})$
2	$\mathbf{y}_2^{(m)} = f_1(\vec{b}_2^{(m)}),\ f_2(\vec{b}_2^{(m)}), \cdots, f_S(\vec{b}_2^{(m)})$
⋮	⋮
N	$\mathbf{y}_N^{(m)} = f_1(\vec{b}_N^{(m)}),\ f_2(\vec{b}_N^{(m)}), \cdots, f_S(\vec{b}_N^{(m)})$

- $f_d(\vec{b}_i^{(m)}) \equiv$ expected response to dose d given the i-th realization of the parameters for model m, d = 1, …, S

- Also,
  - $f_y^{(m)}(y;\ d)$ = posterior density of $y^{(m)}$ (d-th col of $\mathbf{Y}^{(m)}$)
  - $R_y = (y \mid y \in (y_L, y_U))$ = an interval of response values
  - $P(R_y \mid d;\ m) = \int_{y_L}^{y_U} f_y^{(m)}(y;\ d)\,dy \quad \rightarrow \quad P_{post}(y \in R_y \mid d,\ m)$

- Hence, $P(d_j \mid R_y;\ m) = \theta_j\ P(R_y \mid d_j;\ m) / \Sigma_j \theta_j\ P(R_y \mid d_j;\ m)$
  = Posterior probability of dose d_j

Allocation fraction