

Bayesian Model Averaging : Proof of Concept Studies with monotone dose response assumptions

Amy Racine

Bern, 2007

Contents

- Background
- Motivating example
- Technical Remarks
- Some Results
- Conclusions

PoC studies

Small studies with **small number of doses** and **little knowledge** about dose response shape are common (# doses including PBO ≤ 4).

Some of Objectives

- To decide whether the substance warrant the next stage (phase II a) development, this is an **internal decision making**.
- To learn about the relationship between **Y (response)** and **dose (and X covariate)**
- To provide the dose range to be studied in phase II a.
 - To make inference about a quantity **Q**, e.g.
 - the mean of **Y** for a given dose, $Q = \mu_d = E(Y|Dose=d)$
 - evidence that **treatment effect** exceeds threshold **C**, e.g.
 $Q = \Pr(\mu_d - \mu_p > C|Data)$
 - the (predictive) probability that an individual measurement **Y***, eg. observed treatment difference at **in next stage development** exceeds threshold **C**: $Q = \Pr(Y^* > C|Data)$
- Since this implies probabilistic statements, a “statistical model” is needed.

Motivating Example in Proof of Concept (PoC)

- Study design : Paralell groups, Placebo control and 3 doses (Low, Med, High) of a compound
- End point : change from Baseline on VAS (0-100) mm at end of study of a short duration study, i.e. an early surrogate endpoint.
- From experience, the change from baseline is approximately normally distributed and with between patient SD between 25 and 30 mm.
- A viable treatment is expected to reduce VAS by 10 mm in compare to PBO, a good treatment is by 25 mm.
- From substance with similar pharmacological effects and preclinical models, the dose response relationship is expected to be monotone.

PoC Criteria

The quantities of interests are the distribution of the treatment differences $(\mu_j - \mu_p)$

PoC of the compound

$P\{(\mu_3 - \mu_p) \geq 10 \text{ mm} \mid \text{data}\} > A$ e.g. 0.5, the Proof of concept (PoC) is achieved.

PoE of lower doses

$P\{(\mu_j - \mu_p) \geq 10 \text{ mm} \mid \text{data}\} > A$

Our challenge is to evaluate

$P\{(\mu_j - \mu_p) \mid \text{data}\}$, for $j=1,2,3$ under an monotone dose response assumption

 Bayesian Model Averaging

Model Space for

Monotone dose response relationship

$Y_{ij} \sim N(\mu_i, \sigma^2)$, $i = p, d_1, d_2, d_3$ and $j=1, \dots, n_j \leftarrow$ Likelihood

- M_1 $\mu_p = \mu_{d1} = \mu_{d2} = \mu_{d3}$
- M_2 $\mu_p = \mu_{d1} = \mu_{d2} < \mu_{d3}$
- M_3 $\mu_p = \mu_{d1} < \mu_{d2} = \mu_{d3}$
- M_4 $\mu_p = \mu_{d1} < \mu_{d2} < \mu_{d3}$
- M_5 $\mu_p < \mu_{d1} = \mu_{d2} = \mu_{d3}$
- M_6 $\mu_p < \mu_{d1} = \mu_{d2} < \mu_{d3}$
- M_7 $\mu_p < \mu_{d1} < \mu_{d2} = \mu_{d3}$
- M_8 $\mu_p < \mu_{d1} < \mu_{d2} < \mu_{d3}$

Model M_1 represents the null model, whereas M_8 the strictly increasing dose response model.

which is the appropriate model ? \rightarrow Model uncertainty

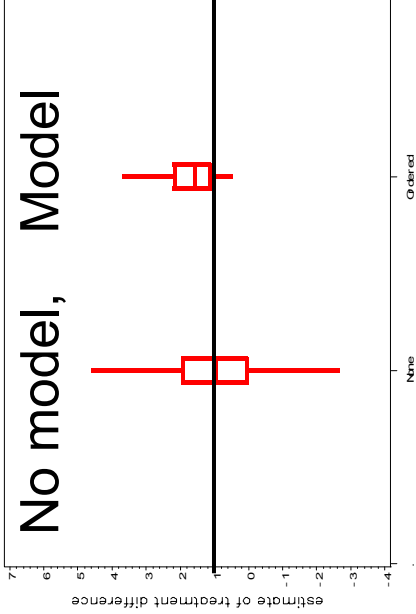
Why use model?

Example 2 doses, each with one measurement & $\sigma^2=1$

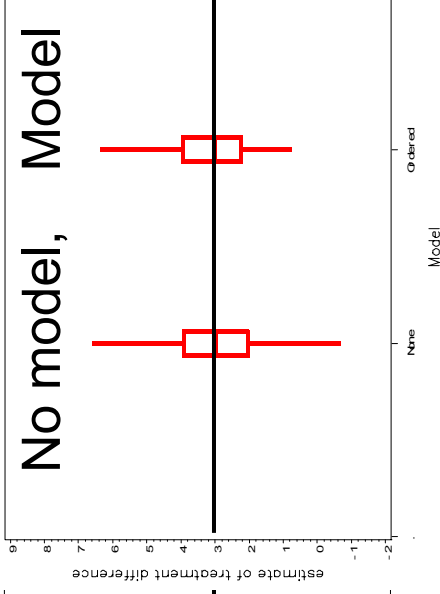
Model $\mu_1 < \mu_2$

Estimate treatment difference from 1000 simulated studies

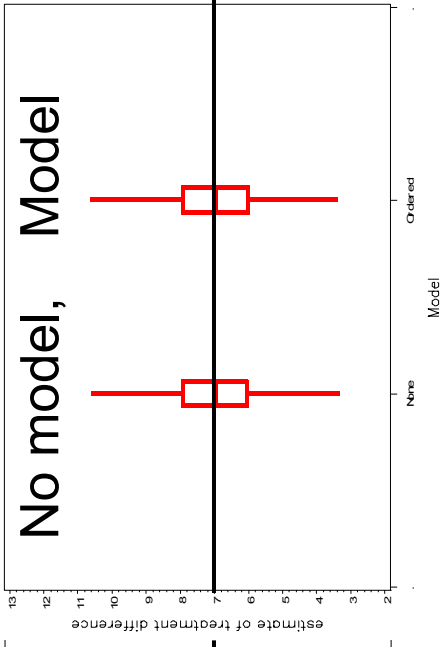
$I=\mu_1 < \mu_2=2$



$I=\mu_1 < \mu_2=4$



$I=\mu_1 < \mu_2=8$



Model approach : decrease variation of the estimate with a bias trade-off

Model Uncertainty (1)

- Model: use $M = (S, \theta)$ for making inferences about Q
 - Structural assumptions S
e.g. link function, heteroscedasticity, conditional independence assumptions, fixed vs. random effects, ...
 - Parameters θ with meaning that is specific to the given model M .
e.g.. regression coefficient and variance when M is a linear model
 - We know how to do inference on θ for a given S . That is, we know how to handle uncertainty of θ given S .
 - But how do we choose S ? How do we handle structural uncertainty?
 - not at all: pre-specify a model S and go with it...
 - examine the data and find a single 'best' S^* .
 - Residual analyses, model selection strategiesThen proceed as if S^* were known, ignore model selection uncertainty.
 - conditional inferences based on S or S^*
- The quantity of interest is Q , not S or S^* ...

Model Uncertainty (2)

- Structural uncertainty is not fully assessed with the S* approach (even after having seen the data, structural uncertainty remains)
- Therefore
 - uncertainty about quantities of interest Q remains.
 - Inferences can be poorly calibrated, i.e.

$$(Q_{\text{est}} - Q_{\text{actual}}) / \text{se}(Q_{\text{est}})$$

is far from $N(0, 1)$.

- If mis-calibration occurs, it is often anticonservative:
confidence interval were not wide enough! Not maintaining type I error.
- If we are concerned about **calibrated inferences**, we have to address **model uncertainty**.

Bayesian Model Averaging : addressing the model uncertainty

- statistical models M_1, \dots, M_J with model parameters $\theta_1, \dots, \theta_J$ (common to “Frequentists”, “Likelihoodists”, “Bayesians”)

$$pr(D|\theta_j, M_j)$$

- 1) Bayesian: priors for parameters of each model

$$pr(\theta_j | M_j)$$

- 2) Bayesian: prior probabilities for model $M_j, j=1, \dots, J,$

$$pr(M_j)$$

- ... what follows (from standard probability calculus)
- the quantity of interest is Q

$$pr(Q|D) = \sum_j pr(Q|M_j, D)pr(M_j|D)$$

a **weighted average** of the posteriors of Q under each model, weighted by the posterior model probabilities

Implementation of Bayesian Model Averaging (2)

- Computational details:
 - Posterior model probabilities
 - The probability of the data for each model
- $$pr(M_j|D) \propto pr(D|M_j) \cdot pr(M_j)$$
- (“marginal likelihoods”, “prior predictive distribution”).
- ... the “computation” of the marginal likelihood is surprisingly difficult

$$pr(D|M_j) = \int pr(D|\theta_j, M_j)pr(\theta_j|M_j)d\theta_j$$

Common computation Methods

- Monte Carlo integration <- sample from the prior
- Monte Carlo integration + Importance sampling
 - Harmonic means <- when the posterior is the proposal density
- Approximations, e.g. using function of BIC

Prior Model Weight

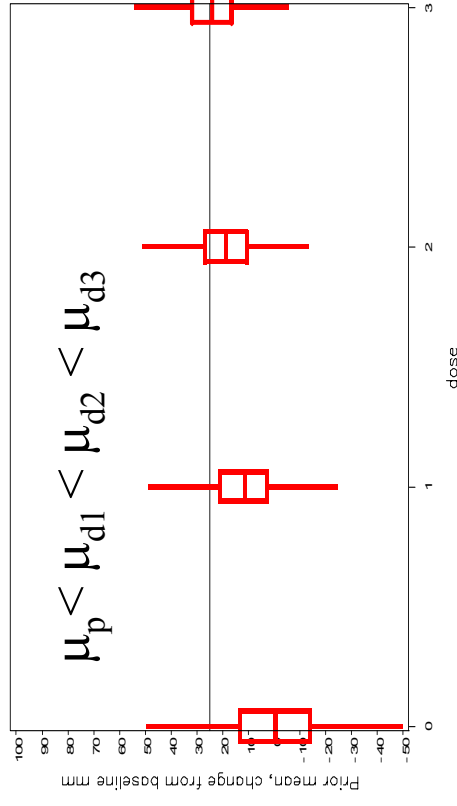
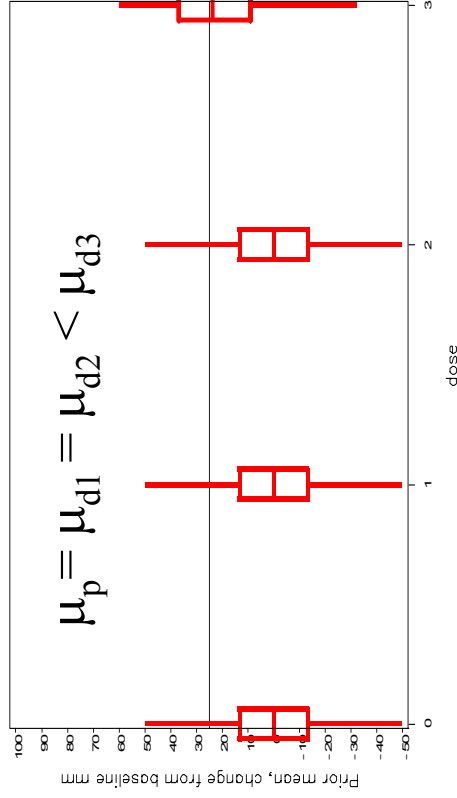
- Example 1
 - All the 8 Models are equally likely, i.e. $w_i=1/8$
- Example 2
 - It is equally likely that the substance is ineffective as it is effective, i.e. $w_1=1/2$, $w_j = 1/14$ for $j=2, \dots$
- Example 3
 - Model 6 is not possible, $w_6=0$
 -

In PoC, we are usually skeptical, the prior weight on null model should be higher than the other 7 models..

Prior on Model Parameters

(mean μ , variance σ^2)

- Under M_1 Null model, $\mu_{d1} \sim N(0, \tau^2)I(a,b)$, a and b are chosen in such a way that it makes sense that is within context of the application, eg. 20, -50, +50.
- Under the other 7 models, the prior is such that the prior on
 - Placebo is same as M_1 , $\mu_p \sim N(0, \tau^2)I(a,b)$
 - The difference of the high dose group to placebo is with mean δ mm, where is δ chosen in the range of the best competitor and the range is within application context



Prior on σ^2 inverted χ^2 with ν df & expectation $\nu\lambda_0$

Implementation method (I)

1. For each of the 8 models draw M posterior samples using Monte-Carlo Markov Chain of $(\mu_p, \mu_{d1}, \mu_{d2}, \mu_{d3}, \sigma^2)$
2. For each of the 8 models calculate the harmonic mean of the posterior samples obtained from step 1.
3. Calculate the posterior model weight of the 8 models which is proportion to the product of their prior weight and the marginal probability of data obtained from step 2.
4. Generates a model $m^{(l)}$ using the posterior model weight in step 3
5. Draw a sample of $(\mu_p, \mu_{d1}, \mu_{d2}, \mu_{d3}, \sigma^2)$ ^(l) from the MCMC samples of model $m^{(l)}$ generated in step 1.
6. Repeat step 4 and 5 M times.
7. Calculate the statistics of interest, eg. distribution of treatment difference for each dose.

Evaluating the performance of the BMA & study design

- Number of patient / dose =20
- Underlying between patient variation = 28 mm
- Scenarios: Number of studies simulated / scenario 1000

scenario	PBO mean	dose 1 mean	dose 2 mean	dose 3 mean
0	5	5	5	5
1	5	5	5	25
2	5	5	25	25
3	5	10	20	25
4	5	25	25	25
5	5	15	15	15
6	5	18	23	25
7	5	15	22	25
8	5	6	23	25
9	5	10	25	15

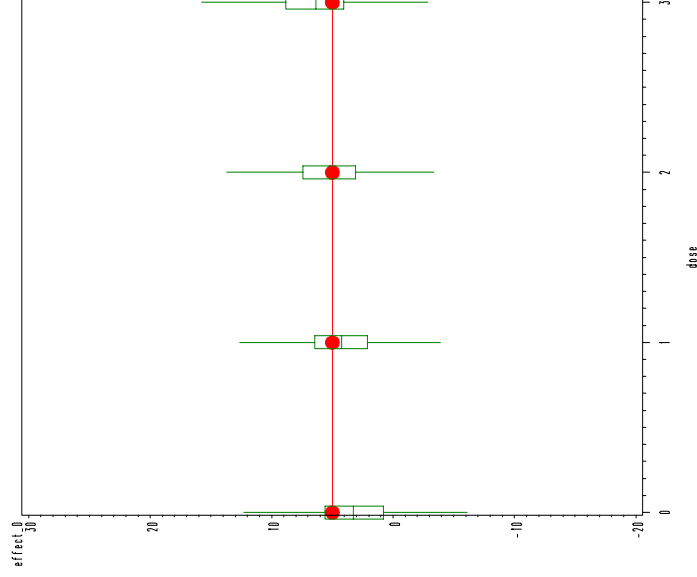
Comparative method:

PoC achieved when the observed treatment difference high dose to placebo $\geq 10\text{mm}$ & is significant at a one sided 5% level.

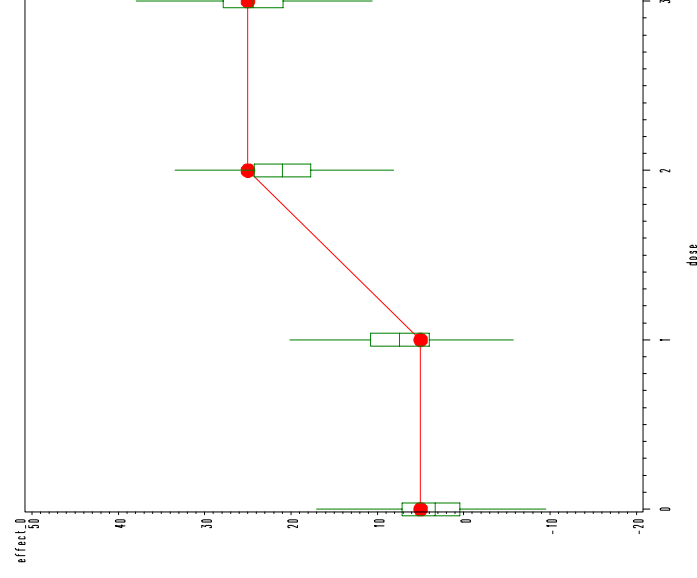
PoE for the lower dose group when the observed treatment difference $> 10\text{ mm}$

Some results from 1000 simulated studies: Prior weight null model 1/2

Box Plot of posterior means of dose response, the underlying scenario



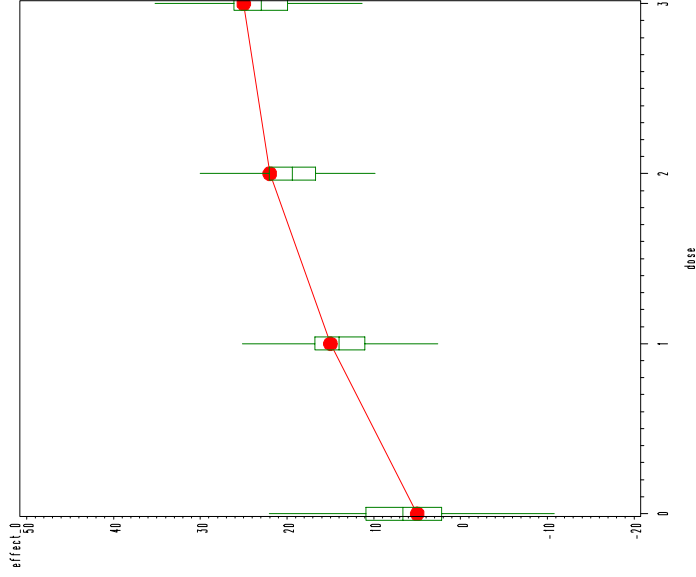
Method	dose 1 PoE	dose 2 PoE	dose 3 PoC
BMA	0.008	0.027	0.063
Comparative	0.010	0.023	0.050



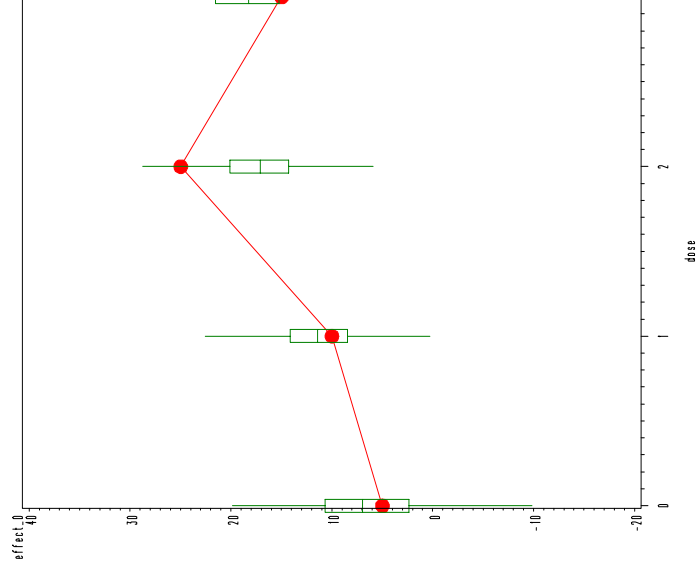
Method	dose 1 PoE	dose 2 PoE	dose 3 PoC
BMA	0.064	0.811	0.893
Comparative	0.122	0.684	0.733

Some results from 1000 simulated studies: Prior weight null model 1/2

Box Plot of posterior means of dose response, the underlying scenario



Method	dose 1 PoE	dose 2 PoE	dose 3 PoC
BMA	0.256	0.619	0.752
Comparative	0.385	0.611	0.745



Method	dose 1 PoE	dose 2 PoE	dose 3 PoC
BMA	0.136	0.475	0.523
Comparative	0.152	0.303	0.310

Summaries of results

- The performances of BMA are either similar to those of comparative procedures or superior.
- It is quite noticeable for the last scenario in which the order constraint does not hold, the performance of BMA procedures are remarkably better than those of the hierarchical testing.
- For small data set as in PoC different **prior weight** leads to a different operating characteristics of the decision criteria, one therefore needs to **calibrate** these prior weight over a wide range of scenarios to assure alignment with our internal requirements.
- Different priors of the model parameters have little impact on the operating characteristics of the decision criteria.
- For small PoC studies with only few doses, this is a good compromise between full parametric model (or models) and no model at all on the dose response relationship.

Extensions

- This BMA approach can easily be implemented as in one go, using **latent variable approach** in WINBUGS. The latent variable here is the model. -> more elegant and efficient.
- Method and implementation can be extended without much effort to **Binary** and **Poisson** outcome variables
- Incorporation of dose independent **covariates** is straight forward

References

- Draper, D. (1995) Assessment and Propagation of Model Uncertainty (with discussion). *JRSS(B)*, 57(1):45-97.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T. (1999) Bayesian model averaging: a tutorial (with discussion). *Statistical Science* 14(4):382-417)
- Kass, R.E, Raftery A.E.
Bayes Factors (1995) *JASA* 90(430) 773-795
- Ting, N. (2006).
Dose Finding in Drug Development. Springer.

Acknowledgments

Novartis

- David Ohlssen
- Michael Branson
- Beat Neuenschwander
- Heinz Schmidli

Academia

- Tony O'Hagan
- Adrian Raftery