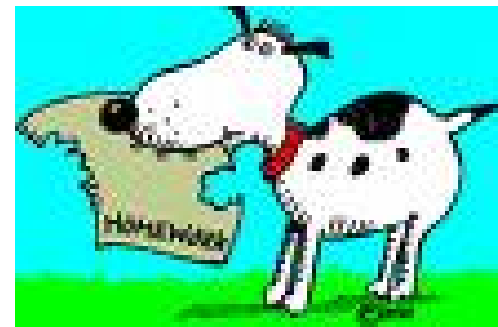
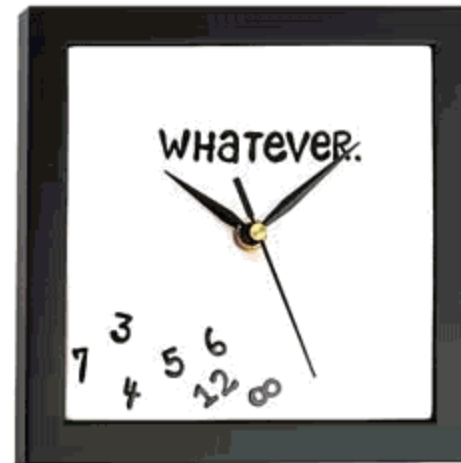
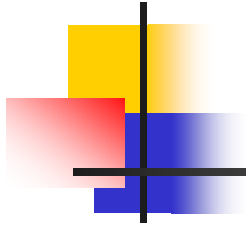


Obnoxious lateness humor



"He's always late - he says he doesn't like to be kept waiting."





Using Bayesian Model Averaging For Addressing Model Uncertainty in Environmental Risk Assessment

Louise Ryan and Melissa Whitney
Department of Biostatistics
Harvard School of Public Health

Resources for the Future
October 22-23, 2007



Motivation

Dose response analysis of arsenic in drinking water showed considerable model sensitivity

— Definition from Breslow and Day (17). ~1975–1980.

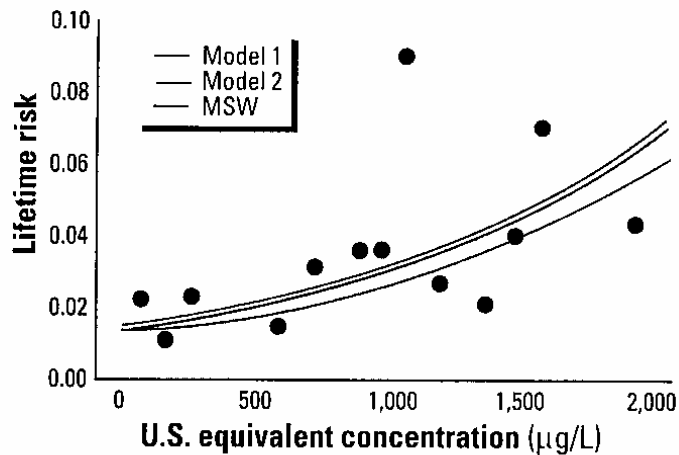


Figure 1. Estimated lifetime death risk for male bladder cancer without comparison population. For a description of models, see Table 6.

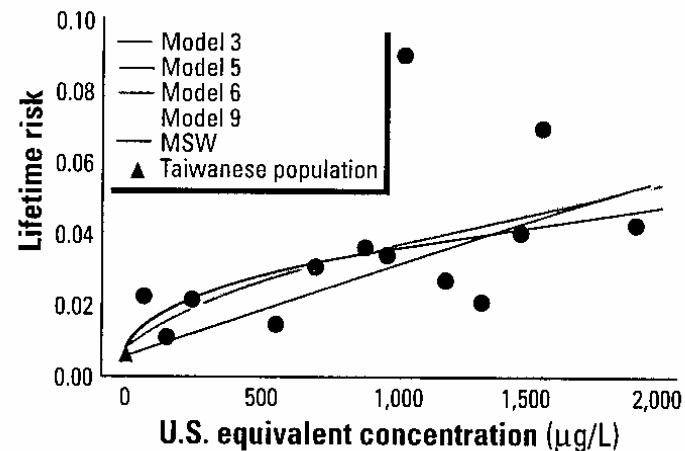


Figure 2. Estimated lifetime death risk for male bladder cancer with Taiwanese-wide comparison population. For a description of models, see Table 6.



Bayesian Model Averaging

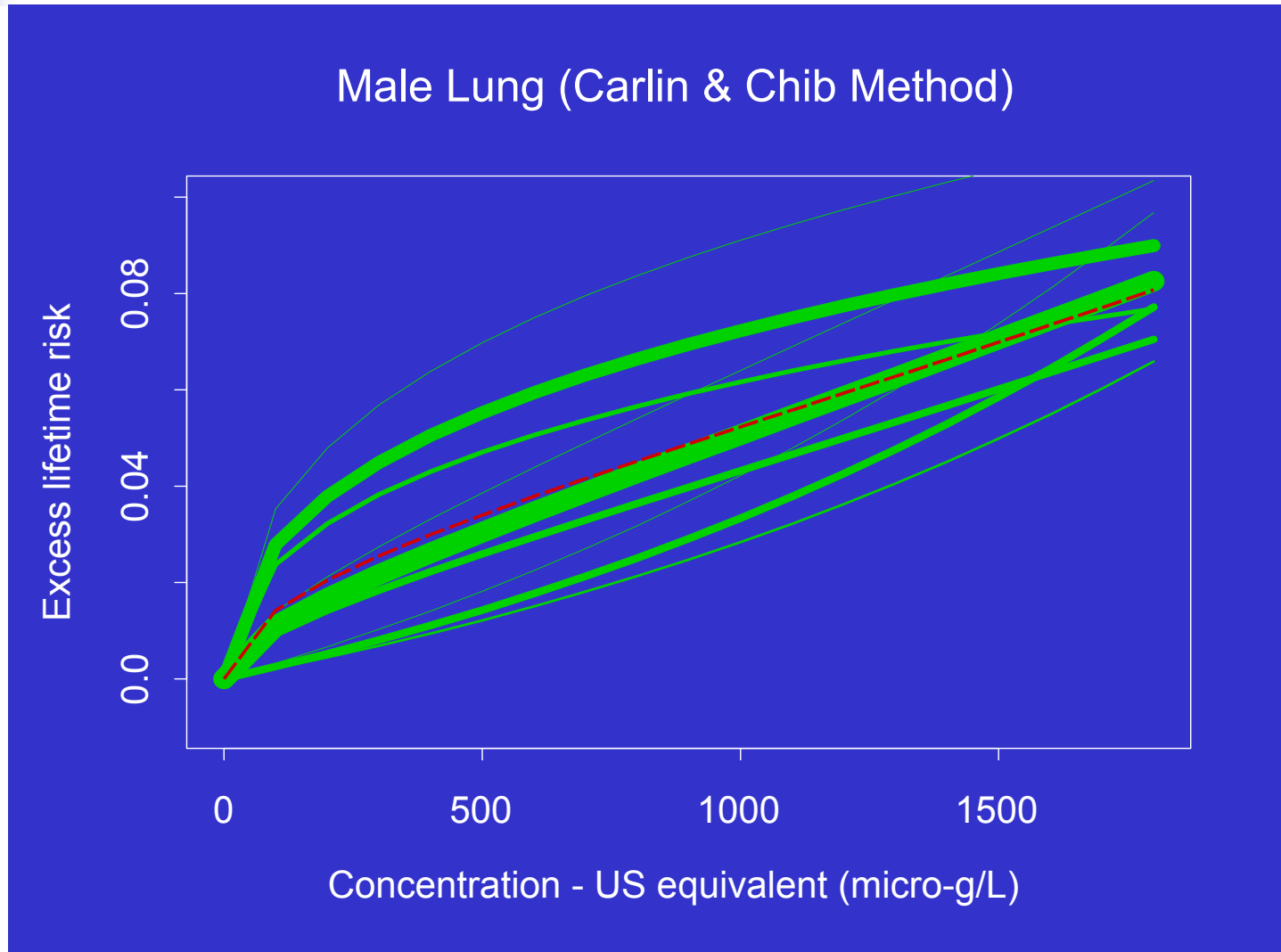
A great way to quantify impact of model uncertainty

- **Down-weights poorly fitting models**
- **Provides a range of risk estimates reflecting statistical variation AND model uncertainty**

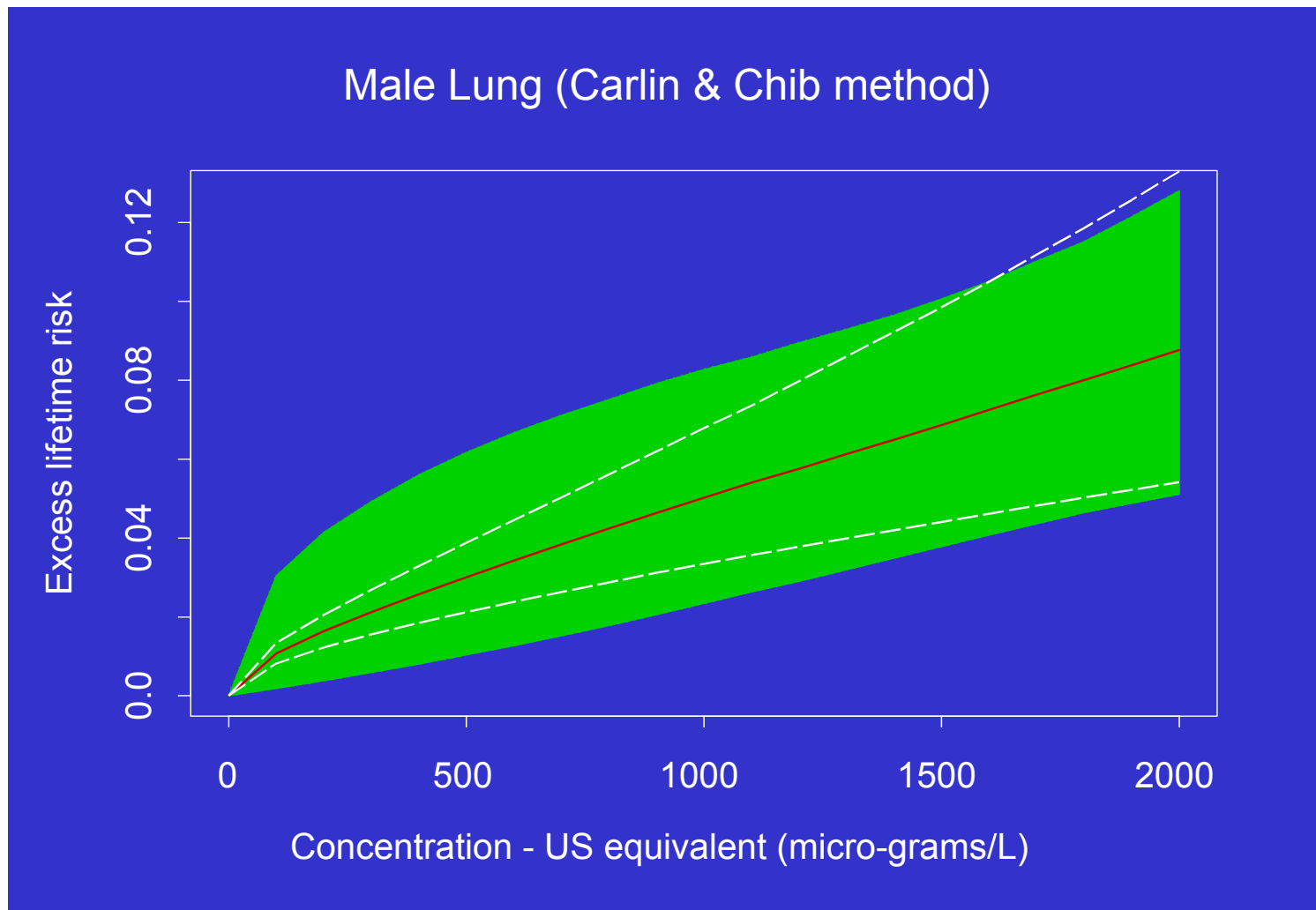
<u>Individual models</u>	<u>BMD (95%CI)</u>
Add, linear dose	42 (40,43)
Multi, linear dose	91 (78,107)
Multi, log dose	70 (60,83)
Model averaged	60 (47, 74)

Model averaged estimate reflects appropriate variation

Fitted curves with thickness proportional to posterior weights



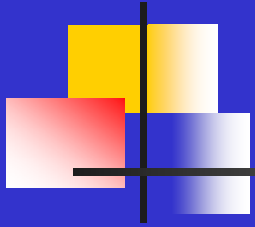
Model Averaged estimate





Basic idea

- Specify a set of suitable models, usually weighted equally a priori
- Compute “posterior” model probabilities, reflecting the likelihood that a model holds, given the observed data
- Average the results with respect to these posterior probabilities – “bad” models get down-weighted, several “good” models can contribute.



Posterior Model Probabilities

Consider models $M=1, \dots, K$. The PMP for model k is

$$P(M=k|D) = \frac{P(D|M=k) P(M=k)}{\sum_{j=1}^K P(D|M=j) P(M=j)}$$

where

$$P(D|M=k) = \int P(D|\theta_k, M=k) P(\theta_k|M=k) d\theta_k$$

This last piece can be a challenge to compute



Calculating Posterior Model Probabilities for Model Averaging

- Fully Bayesian Methods:
 - Closed form solutions (rarely exist for biological models)
 - MCMC methods, for example:
 - Carlin and Chib Method
 - Reverse Jump MCMC
 - Stochastic Search Variable Selection
 - Gibbs Variable Selection
- Frequentist Approximations, e.g. BIC Approx:

$$BIC = 2 \log pr(D | \hat{\theta}_k, M_k) - (p_k / 2) \log(n)$$

$$p(M_k | D) = \exp(0.5 BIC_k) / \sum \exp(0.5 BIC_i)$$



Uncertainty Estimation in BMA

After obtaining posterior probabilities and using classical procedures to obtain any quantity of interest, Δ , (for example BMD) we can then use model weights to obtain the BMA estimate of their *unconditional* expectation and variance

Model selection uncertainty component

$$\hat{\Delta}_{BMA} = \sum_k \hat{\Delta}_k \hat{p}(M_k | D)$$

$$\hat{V}_{BMA}(\hat{\Delta}) = \underbrace{\sum_k \hat{V}[\hat{\Delta} | M_k] \hat{p}(M_k | D)}_{\text{Estimated via a classical procedure or bootstrap}} + \sum_k (\hat{\Delta}_k - \hat{\Delta}_{BMA})^2 \hat{p}(M_k | D)$$

Estimated via a classical procedure or bootstrap



Dataset 1: Uncertainty Due to Choice of Link Function/Dose-Response Model

Dose	Number of Subjects	Number of Responses
0	50	1
21	49	15
60	45	20

Models for Dataset 1

$$1: \frac{e^{\beta_1 + \beta_2 \log(d)}}{1 + e^{\beta_1 + \beta_2 \log(d)}}$$

$$2: \Phi(\beta_1 + \beta_2 \log(d))$$

$$3: 1 - e^{-\beta_1 d - \beta_2 d^2}$$

$$4: 1 - e^{-\beta_1 d^{\beta_2}}$$

$$5: \beta_3 + (1 - \beta_3) \frac{e^{\beta_1 + \beta_2 \log(d)}}{1 + e^{\beta_1 + \beta_2 \log(d)}}$$

$$6: \beta_3 + (1 - \beta_3) \Phi(\beta_1 + \beta_2 \log(d))$$

$$7: \beta_3 + (1 - \beta_3) \left(1 - e^{-\beta_1 d - \beta_2 d^2}\right)$$

$$8: \beta_3 + (1 - \beta_3) \left(1 - e^{-\beta_1 d^{\beta_2}}\right)$$

$$9: \frac{e^{\beta_1 + \beta_2 \log(d + \beta_3)}}{1 + e^{\beta_1 + \beta_2 \log(d + \beta_3)}}$$

$$10: \Phi(\beta_1 + \beta_2 \log(d + \beta_3))$$

$$11: \left(1 - e^{-\beta_1 (d + \beta_3) - \beta_2 (d + \beta_3)^2}\right)$$

$$12: \left(1 - e^{-\beta_1 (d + \beta_3)^{\beta_2}}\right)$$

$$13: \frac{e^{\beta_1 + \beta_2 d}}{1 + e^{\beta_1 + \beta_2 d}}$$

$$14: \Phi(\beta_1 + \beta_2 d)$$

Zero background models

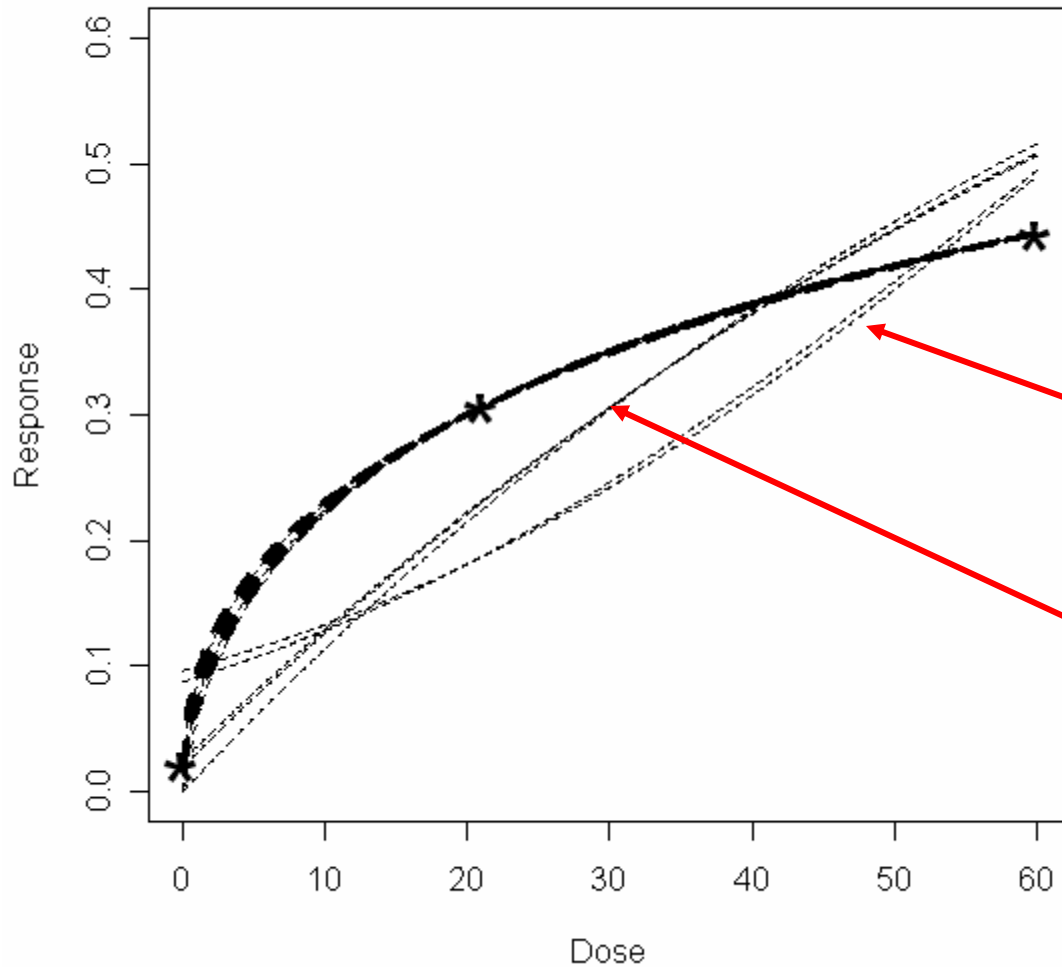


Notes on model fitting

- Fitted models using nlminb in R
- For models 4 through 8, the mle of β_3 is simply r/n
- Benchmark dose is solution to
$$p(x) - p(0) = \text{bmr}$$
(we used $\text{bmr} = 10\%$)
- Lower limits computed using parametric bootstrap

Results

Dose response for dataset 1



**Zero background models
have 0 posterior
probability**

**Models 13 and 14
(posterior probs ~2%)**

**Models 4, 8 and 12
(posterior probs ~5%)**

**Six other models have
posterior prob of ~14%**

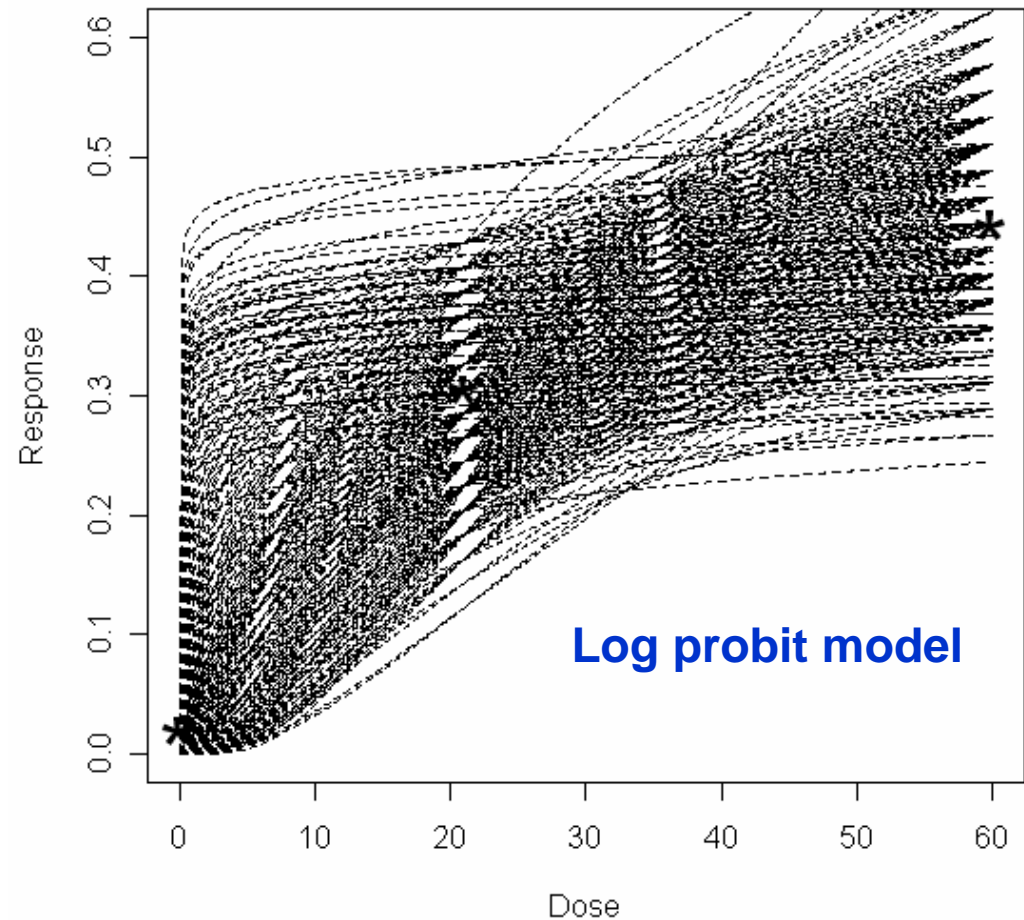
Benchmark Dose estimates (bmd) and associated lower limits (bmdl)

Model	pmp	bmd	bmdl	
[1,]	0.000	1.842	0.000	} Zero background models
[2,]	0.000	2.292	0.000	
[3,]	0.000	1.345	0.000	
[4,]	0.000	8.708	0.000	
[5,]	0.144	2.340	0.000	Model averaged BMD =3.91 BMDL=0.0238
[6,]	0.144	2.842	0.000	
[7,]	0.144	1.791	0.000	
[8,]	0.047	9.362	7.394	
[9,]	0.144	2.579	0.000	
[10,]	0.144	2.929	0.000	
[11,]	0.144	2.034	0.000	
[12,]	0.048	9.542	7.552	} Simple empirical models
[13,]	0.017	22.639	19.429	
[14,]	0.025	20.970	18.187	

Why are the BMDLs so low?

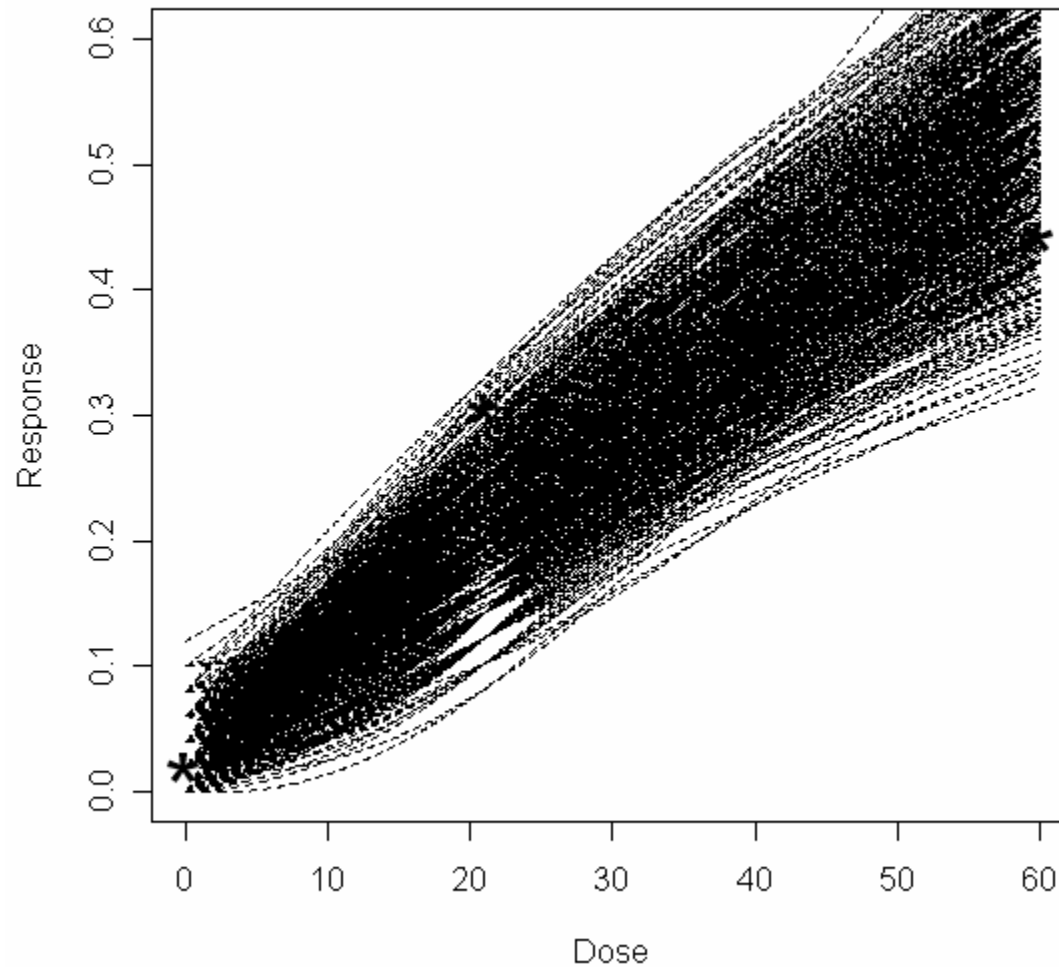
- Bootstrap can easily generate non-monotonic patterns. Zero background models fit weirdly

Dose response for dataset 1



Multistage much more stable

Dose response for dataset 1, model 8





Some comments

- There are many problems with BMD calculations for small simple datasets like dataset 1. Model averaging alone can't fix these problems! Need biological input
- REAL value of BMA is the potential to use informative priors. More later



Dataset #2 – Uncertainty Due to Covariate (Strata) Adjustment

Issue for Frambozadrine Dataset:

- Should we combine males and females?
- Do we need separate dose/response curves?



Dataset 2: Frambozadrine

	Dose(mg/kg-day)	Total no rats	Hyperkeratosis
Male			
	0	47	2
	1.2	45	6
	15	44	4
	82	47	24
Female			
	0	48	3
	1.8	49	5
	21	47	3
	109	48	33

Approaches:

1. Fit two separate models
2. Pool data
3. Multivariate Logistic Regression



A. Naïve Analysis

Model (Logistic Regression)	AIC
$\text{Pr}(\text{response}) = g(-2.50124 + 0.030856 * \text{dose} - 0.229133 * \text{sex} + 0.001060 * \text{dose} * \text{sex})$	40.02
$\text{Pr}(\text{response}) = g(-2.53102 + 0.03151 * \text{dose} - 0.17515 * \text{sex})$	38.04
$\text{Pr}(\text{response}) = g(-2.6389 + 0.03126 * \text{dose})$	36.37



Limitations to Traditional Approach

- Researchers consider a variety of plausible models, but then select and draw inferences from only single “final” model, ignoring other plausible alternatives
- Underestimates true variability and uncertainty due to model selection process
- Results in over-confident decision-making (Draper, 1995)



Bayesian Framework for Model Selection (GVS)

- Likelihood:

- $Y[i] \sim \text{binom}(p[i], n[i])$
- $\text{logit}(p[i]) = \beta_0 + \beta_1 * \text{dose} * \gamma_1 + \beta_2 * \text{sex} * \gamma_2 + \beta_{12} * \text{dose} * \text{sex} * \gamma_{12}$

- Priors:

- $\gamma_i \sim \text{dbern}(0.5)$ where $\gamma=1$ means this variable is included in the model
- $\beta_i \sim \text{norm}(0, \text{tau})$
- $\text{tau} \sim \text{dgamma}(0.1, 0.1)$

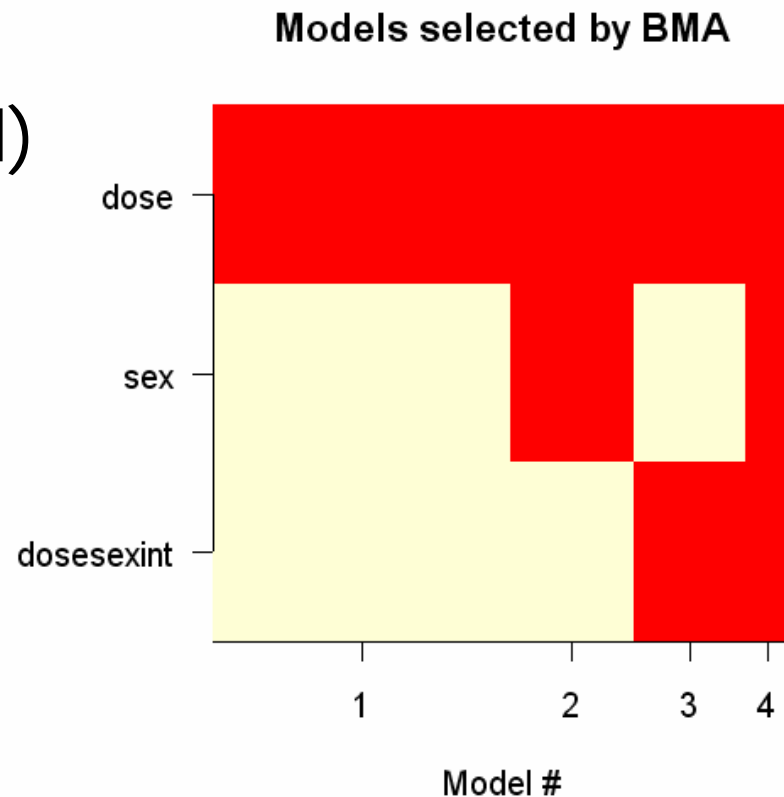


WinBUGS Output, GVS

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	-2.626	0.2417	0.005746	-3.104	-2.617	-2.153	501	1000
beta1	0.03173	0.003439	9.991E-5	0.02498	0.03176	0.03839	501	1000
beta12	-0.002485	0.2493	0.008212	-0.5068	0.002677	0.4812	501	1000
beta2	-0.0311	0.2248	0.006597	-0.4681	-0.03721	0.4426	501	1000
deviance	34.54	2.056	0.06515	32.33	34.0	40.11	501	1000
g[1]	1.0	0.0	2.236E-12	1.0	1.0	1.0	501	1000
g[2]	0.454	0.4979	0.01377	0.0	0.0	1.0	501	1000
g[3]	0.011	0.1043	0.002918	0.0	0.0	0.0	501	1000
p[1]	0.06906	0.01554	3.694E-4	0.04294	0.06804	0.1041	501	1000
p[2]	0.0715	0.01585	3.756E-4	0.04481	0.07038	0.1072	501	1000
p[3]	0.1059	0.01955	4.543E-4	0.07107	0.1054	0.1493	501	1000
p[4]	0.494	0.04923	0.001505	0.3986	0.4946	0.5917	501	1000
p[5]	0.06745	0.01582	3.994E-4	0.04132	0.06614	0.1042	501	1000
p[6]	0.07105	0.01629	4.096E-4	0.04385	0.06991	0.1087	501	1000
p[7]	0.1221	0.02184	5.296E-4	0.08406	0.1206	0.1706	501	1000
p[8]	0.6877	0.05459	0.001613	0.5777	0.6883	0.7918	501	1000
pmdl[1]	0.0	0.0	2.236E-12	0.0	0.0	0.0	501	1000
pmdl[2]	0.546	0.4979	0.01377	0.0	1.0	1.0	501	1000
pmdl[3]	0.0	0.0	2.236E-12	0.0	0.0	0.0	501	1000
pmdl[4]	0.443	0.4967	0.0145	0.0	0.0	1.0	501	1000
pmdl[5]	0.0	0.0	2.236E-12	0.0	0.0	0.0	501	1000
pmdl[6]	0.0	0.0	2.236E-12	0.0	0.0	0.0	501	1000
pmdl[7]	0.0	0.0	2.236E-12	0.0	0.0	0.0	501	1000
pmdl[8]	0.011	0.1043	0.002918	0.0	0.0	0.0	501	1000

C. Bayesian Model Averaging Using BIC Approximation to Find Posterior Model Probabilities

- Implemented using
Volinsky's R package
BMA (function BIC.GLM)





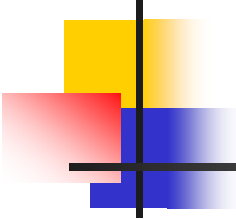
BIC Approximation, bic.glm Output

	p!=0	EV	SD	model 1	model 2	model 3	model 4
Intercept	100	-2.5826932	0.250768	-2.603888	-2.531016	-2.616012	-2.501244
dose	100.0	0.0314884	0.003892	0.031256	0.031509	0.032337	0.030856
sex	29.2	-0.0552155	0.211607	.	-0.175148	.	-0.229133
dosesexint	27.0	-0.0002177	0.002886	.	.	-0.001550	0.001060
nVar				1	2	2	3
BIC				-7.084206	-5.333404	-5.119532	-3.276673
post prob				0.515	0.215	0.193	0.077



Comparison: All 3 Methods

Variable	Traditional	Fully Bayesian	BIC Approx.
Dose	0.03125 (0.00339)	0.03173 (0.003439)	0.031488 (0.003892)
Sex	-----	-0.0311 (0.2248)	-0.055216 (0.211607)
Dose*Sex	-----	-0.002485 (0.2493)	-0.000218 (0.002886)



Fully Bayesian vs. BIC Approx. Posterior Probabilities

Variable	Fully Bayesian	BIC Approx.
Dose	Pr(dose) = 1.0 EV: 0.03173 SD: 0.003439	Pr(dose) = 1.0 EV: 0.03149 SD: 0.003892
Sex	Pr(sex) = 0.454 EV: -0.0311 SD: 0.2249	Pr(sex) = 0.292 EV: -0.05522 SD: 0.211607
Dose*Sex	Pr(dxs): 0.011 EV: -0.002485 SD: 0.2493	Pr(dxs) = 0.27 EV: -0.0002177 SD: 0.002886

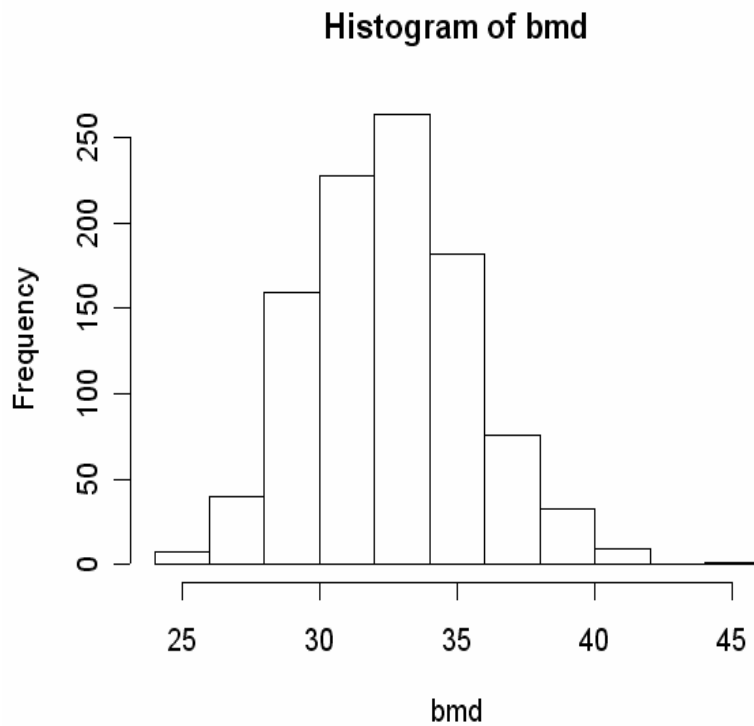


Comparison of Approaches

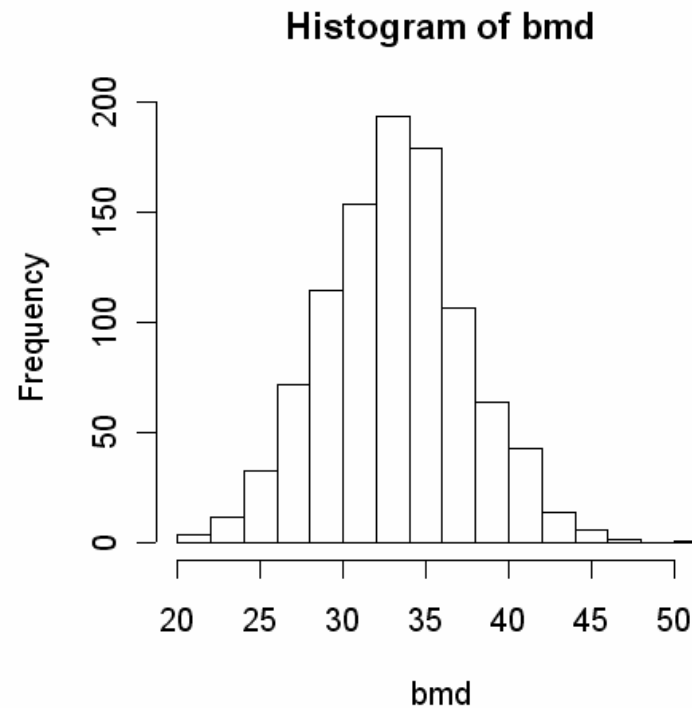
- Traditional Approach: Widely Adopted, Easily Implemented
- BIC Approach: Easily implemented, properties well-understood; Strong penalty for inclusion of covariates → favors simple models (not ideal from toxicologist's perspective)
- Fully Bayesian Approach: Most intensive to implement, may not be as stable for some models, but other fully Bayesian approaches available as well [Carlin and Chib, Closed Form Solution (in a few cases), RJ MCMC, GVS, SSVS]

BMDS: Histograms

■ Males



■ Females





Conclusions

- BMA helps avoid problem of choosing one model (or one dataset) and does appealing synthesis over a class of feasible models
- NOT a panacea! Only as good as the models it averages over
- The kind of uncertainty it captures is limited
- Real potential is to average over models that reflect biology and/or measurable sources of heterogeneity



A proposal

- Form compound-specific expert panels to identify model classes and combine using BMA
- Since using a Bayesian framework anyway, can include some Bayesian models in the model class
- Can include multiple studies, species, gender with careful and appropriate modeling
- Concept: synthesize all relevant information to inform reliable decision making