Risk estimation, model uncertainty and risk assessment practice

A. John Bailer (Miami U) and Matthew Wheeler (UNC/NIOSH)

Contact information:
A. John Bailer
Department of Statistics
Miami University
Oxford, Ohio 45056
T: 513.529.7828 (On your keypad 7828=?)
baileraj@muohio.edu
www.users.muohio.edu/baileraj

Collaborators: Robert Noble (Miami U./GSK), Robert Park (NIOSH)

Presented at ICRA4 in Limassol, Cyprus on 28 May 2011

Outline

1. Motivation
2. Initial interest and where to begin
3. Where might in MA fit in QRA and how might it be implemented?
4. Does it work?
5. Can it be improved?
6. Future

First, and most importantly,

THANK YOU

To the organizing committee and our outstanding hosts …

To all of the participants in the Conference …

1. Motivation – quantitative risk assessment (QRA)/risk estimation

* Health-related QRA: history - DBRA @ NIEHS; REB@NIOSH and SEG(NRC) [bonus points for acronym ID!]
* Hazard identification + exposure assessment + dose-response modeling + risk characterization (NRC reports)
* QRA: start with relationship between hazard exposure level and a particular adverse response (based on animal or human data), say a dose-response (D-R) relationship
* QRA: Invert this D-R relationship to predict the exposure level associated with some specified response level yields endpoints such as benchmark dose/concentration
* [apologies: use “dose” and “concentration” interchangeably - need to be careful if working in a tox. PBPK/PBTK context]

Data sources used to derive risk estimates?

Animal (toxicology) data: rodents randomly assigned to dose groups

- Responses – often dichotomous (e.g. tumor presence)
- Predictor variable(s): dose term(s)

One familiar source of QRA uncertainty: animal-to-human extrapolation.

Epidemiology data: Observational studies (e.g. retrospective cohort studies)

- Responses – any measurement scale
- Predictor variables: exposure of interest (re-constructed), confounders, effect modifiers, shape of exposure estimate

Another familiar source of QRA uncertainty: exposure assessments in studies involving humans (mixture issues as well).

(initial focus on QRA based on animal studies …)

Dose-response models?

E[Response] = *f*(dose) + distributional assumptions

Example: animal carcinogenicity studies

Response = number of tumor-bearing animals

$E[\text{Response}] = n \{ 1 - \exp(-\theta_0 - \theta_1 d - \theta_2 d^2) \} = n \pi(d)$

("quantal multistage model")

Risk endpoints?

BMD: d s.t. $BMR = [\pi(d) - \pi(0)]/[1 - \pi(0)]$ = extra risk(ER) at d

(Crump proposed alternative to NOEL/NOEC)

- Alternatively, could fix dose and estimate an ER/AR

QRA uncertainty: BMD or ER/AR may differ dramatically as a function of the model selected.

Strategies:

a. Pick “best fitting” model and use it for all risk estimates.
   [e.g. “best fit” = largest P-value in X^2 GOF test]

b. Select the model yielding the smallest BMD estimate – a
   “conservative” option.

c. Fit a number of models and report a range of estimates.

d. Something else [today’s presentation!].

2. Initial interest and where to begin

* Former student working in a risk research group at NIOSH (Matt Wheeler)
* USEPA BMD software publicly available and people could fit a number of D-R models and obtain BMDs for each
* Began serving on SEG committee of the NRC
* Colleague with background in **model averaging (MA)** hired at Miami (Bob Noble)
*  would MA work for QRA?

3. Where might MA fit in QRA and how might it be implemented?

Q: Where would MA incorporated in QRA?

A: ?

Q: How might MA help in the QRA process?

A: ?

Example 1 – dose terms in same quantal model

* g non-zero concentration conditions (d_1, …, d_g) along with a control condition (d_0=0)

* number of organisms exposed to the i^{th} condition is n_i.

* number of organisms developing tumors is recorded, say Y_i for the i^{th} concentration condition

* π_i =Pr(developing a tumor given exposure to hazard concentration d_i)

* Assume: data are "generated" in a binomial expt,

i.e. $Y_i \sim \text{Bin}(n_i, \pi_i)$

[Aside: models incorporating over-dispersion patterns relative to the binomial also could be considered]

* Quantal multistage model:

$$\pi_i = 1 - \exp(-\theta_0 - \theta_1 d_i - \theta_2 d_i^2 - \theta_3 d_i^3 - ... - \theta_k d_i^k)$$

where $k \leq g$ & $\theta_i \geq 0$.

* Extra risk $\quad ER(d_i) = \dfrac{\pi_i - \pi_0}{1 - \pi_0}$

* For low concentrations close to zero, $ER(d) \approx \theta_1 d$.

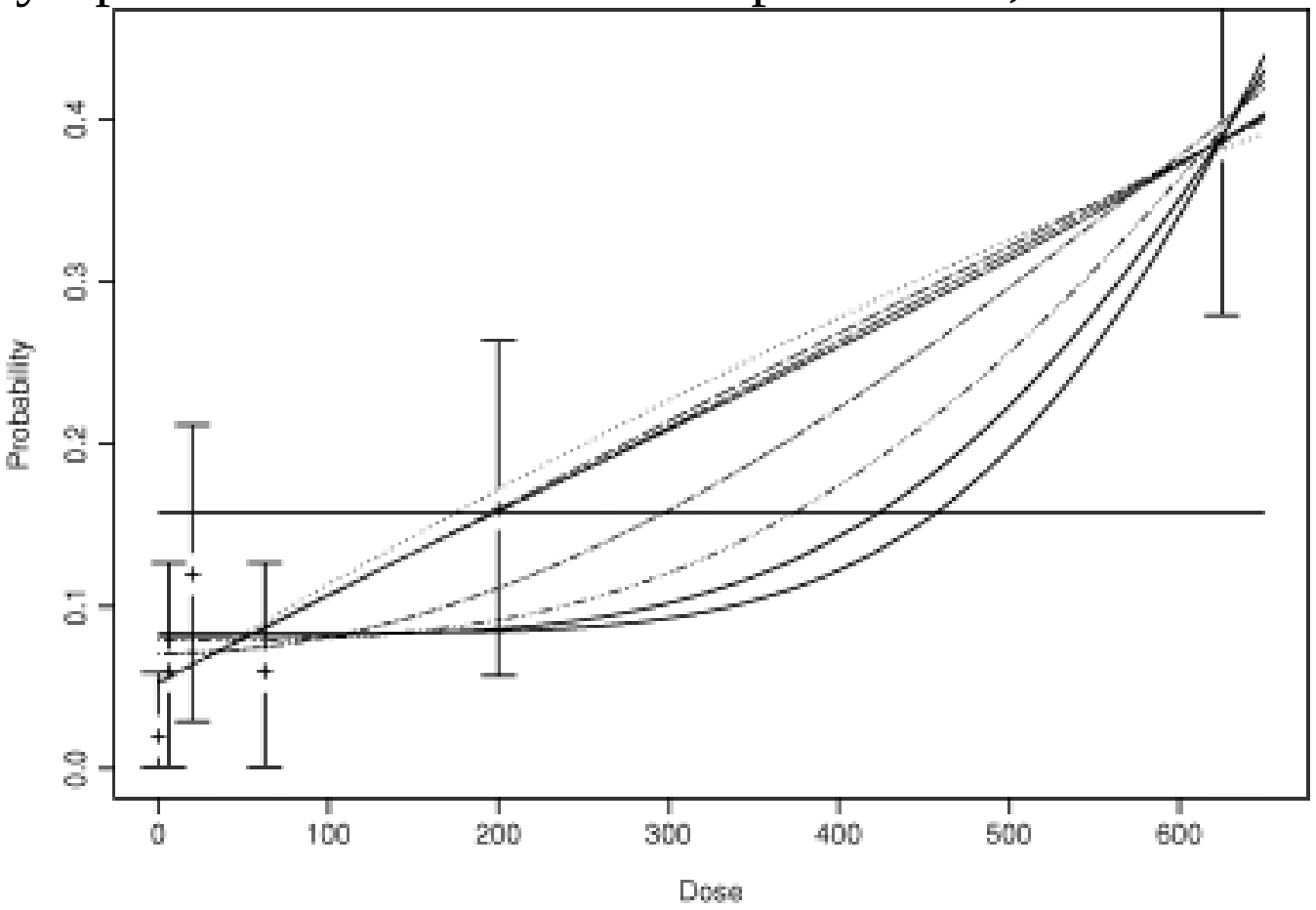
Example 1 – data illustration

Data (NTP 1993) lymphocytic malignant lymphomas in FM exposed to 1,3-butadiene

Conc. (ppm)	*0*	*6.25*	*20*	*62.5*	*200*	*625*
y_i/n_i	1/50 (2%)	3/50 (6%)	6/50 (12%)	3/50 (6%)	8/50 (16%)	31/80 (39%)
surv-adj	2.2%	8.5%,	19.6%	13.5%	37.8%	70.5%

* Poly-3 adjusted proportion as suggested by Bailer and Portier (1988)]

Figure: Plot of the estimated probability of lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene (NTP 1983).



Example 1 Issue:

If we restrict the analysis to a multistage model, then how many dose terms?

Note: most/all of the model fits satisfied a GOF test

What if we didn't have to choose only one?

Example 1 Question:

Q1: Form of the dose effect in dose response model?

Early thinking: can we implement MA with output from standard packages (e.g. USEPA BMDS)?

Some background …

Bayesian Model Averaging (BMA) formally incorporates model uncertainty in an analysis

* family of models is defined, say M

* data are collected

* posterior probabilities (or model weights) associated with the models are derived

* summaries of BMA can be found in a number of papers including Raftery et al. (1997) and Hoeting et al. (1999)

* distribution of some characteristic of a model, e.g. excess risk or BMD or BMDL, is derived over some space of possible models.

* posterior distribution of this characteristic given the data is

$$pr(\Delta \mid D) = \sum_{m=1}^{N_m} pr(\Delta \mid M_m, D) pr(M_m \mid D)$$ where

$pr(M_m \mid D)$ = posterior distribution of the model given the data

N_m= number of possible models considered

* posterior distribution of the model given the data can be obtained using Bayes' Rule as

$$pr(M_m \mid D) = \frac{pr(D \mid M_m) pr(M_m)}{\sum_{l=1}^{N_m} pr(D \mid M_l) pr(M_l)}.$$

* ingredients: *prior probability* of the model, $pr(M_m)$, & the *likelihood of the data given the model*, $pr(D|M_m)$.

* **likelihood** of the data given a particular model is $pr(D|M_m) = \int pr(D|\theta_m, M_m) pr(\theta_m|M_m) d\theta_m$ where

$\boldsymbol{\theta}_m$ =vector of parameters included in model M_m, i.e. θ_0, θ_1, … θ_k.

Given the posterior distribution, relevant summaries can be constructed

* *posterior mean excess risk* at dose d_i can be constructed as $E[\Delta(d)|D] = \sum_{m=1}^{N_m} \Delta_m(d) pr(M_m|D)$ where $\Delta_m(d)$ is the excess risk at dose d_i calculated from model M_m.

Comments:

* model prior: "$pr(M_m)$" often assumed to be $\frac{1}{N_M}$

* parameter prior: "$pr(\theta_m \mid M_m)$" not intrinsically interesting in this application (although may be in a full Bayesian analysis)

* Can employ approx. to determine the posterior model probability that is related to a penalized likelihood –

BIC = -2 ln(L) + p ln(n) (p = number of model parameters)

* Note: The above is an approximation to "$pr(M_m \backslash data)$," a fully Bayesian approach can be computed using MCMC methods. (e.g., reversible jump MCMC)

* BIC provides a close approximation to the Bayes factor when a unit-information prior on the parameter space is used [Kass and Wasserman].

* Raftery evaluated the accuracy of the approximation of BIC within the framework of generalized linear models.

* When the prior model probabilities are equal, the posterior probability of a model given the data is proportional to exp(−0.5 BIC), i.e., $P(M_m|D) = \frac{\exp(-0.5\, BIC_m)}{\sum_{r=1}^{Nm} \exp(-0.5\, BIC_r)}$

Model averaging (MA) used in a wide variety of RA contexts.

* Kang *et al.* looked at MA in microbial risk assessment
* Moon *et al.* extended this work using a different IC.
* Faes *et al.* looked at MA with fractional polynomials being fit to microbial RA data
* Morales *et al.* use a fully Bayesian analysis where priors on both the models considered and on the parameters of each of these models were fully specified in the MA.
* Buckland *et al.* proposed simpler methods, where weights are based upon the penalized likelihood functions formed from the AIC and BIC.

[AIC $=-2\log L+2p$ and the BIC $=-2\log(L)+p\log(n)$, and p is the number of parameters in the model, L is the maximum likelihood value, and n represents the sample size]

Aside:

Standard practice assumes your selected model is true –

$$\Pr(\text{my model} \mid \text{data}) = 1$$

and

$$\Pr(\text{any other model} \mid \text{data}) = 0$$

Is this more palatable?

Model space for the multistage model in Example 1?

Model ID[1]	Term present? d	d^2	d^3	d^4	d^5	Model ID[1]	Term present? d	d^2	d^3	d^4	d^5
1	0	0	0	0	0	33	0	0	0	0	1
3	1	0	0	0	0	35	1	0	0	0	1
5	0	1	0	0	0	37	0	1	0	0	1
7	1	1	0	0	0	39	1	1	0	0	1
9	0	0	1	0	0	41	0	0	1	0	1
11	1	0	1	0	0	43	1	0	1	0	1
13	0	1	1	0	0	45	0	1	1	0	1
15	1	1	1	0	0	47	1	1	1	0	1
17	0	0	0	1	0	49	0	0	0	1	1
19	1	0	0	1	0	51	1	0	0	1	1
21	0	1	0	1	0	53	0	1	0	1	1
23	1	1	0	1	0	55	1	1	0	1	1
25	0	0	1	1	0	57	0	0	1	1	1
27	1	0	1	1	0	59	1	0	1	1	1
29	0	1	1	1	0	61	0	1	1	1	1
31	1	1	1	1	0	63	1	1	1	1	1

[1] Model ID number is based upon terms present in the model. In particular, the
ID # = 1 + 2*"d" + 4*"d^2" + 8*"d^3" + 16*"d^4" + 32*"d^5" where "d^j" = 1 if dose term d^j is present in the model and "d^j" =0 if not (j=2,3,4,5).

Extra risk of tumor development for a given model.

Let $\Delta(d \mid M_j) = d_i$ such that $ER(d_i \mid M_j) = \frac{\pi_{i|j} - \pi_{0|j}}{1 - \pi_{0|j}} = 0.001$

(i.e. the dose estimated to have an ER=0.001 for model M_j).

BMA estimate:

$$\Delta(d) = \sum_{j=0}^{31} \Delta(d \mid M_{2j+1}) pr(M_{2j+1} \mid D)$$

Variance component due to model uncertainty

$$\sigma^2[\Delta(d)] = \sum_{j=0}^{31} \left[\Delta(d \mid M_{2j+1}) - \Delta(d)\right]^2 pr(M_{2j+1} \mid D)$$

Figure 2. Plot of posterior model probability associated with each of the 32 candidate models for multistage models fit to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene [Data source: NTP (1983).].

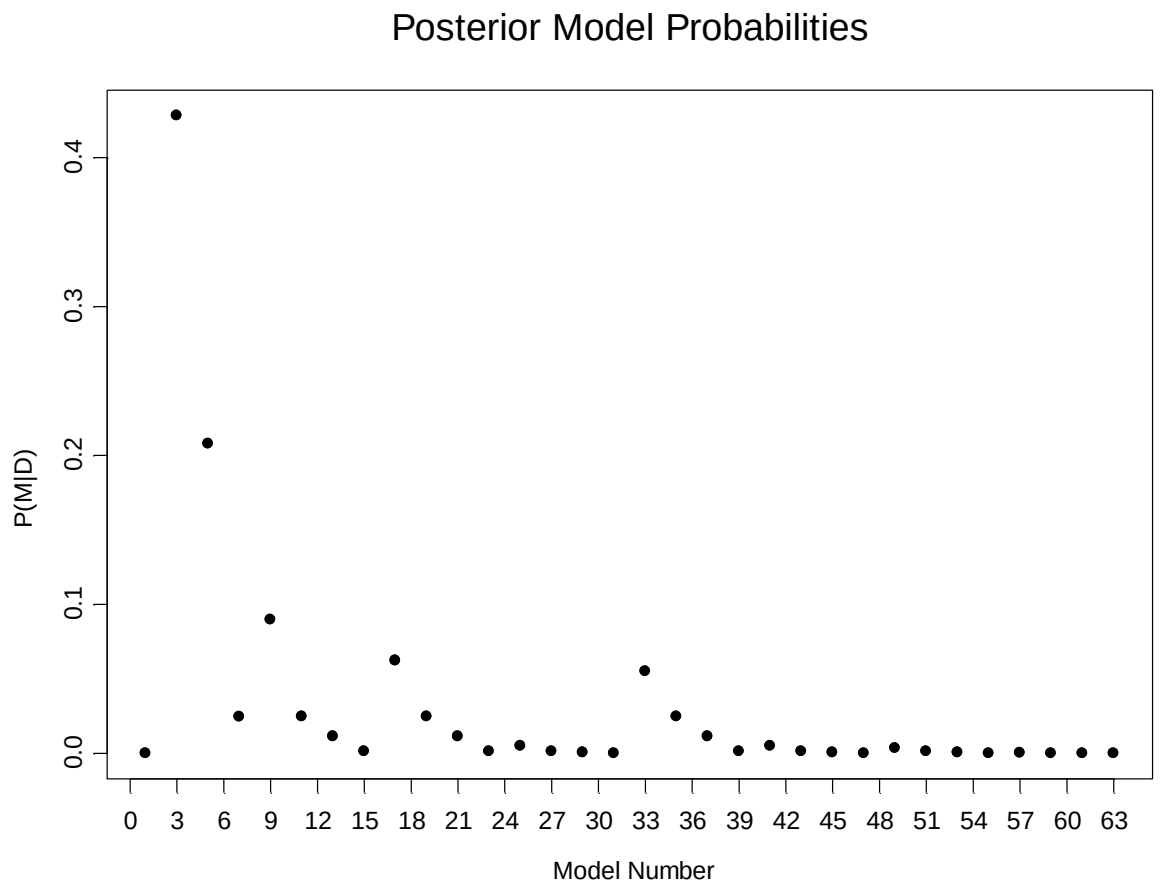


Figure 3. Plot of standard deviation in dose-specific model predictions of tumor probabilities for different multistage models fit to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene [NTP (1983).].

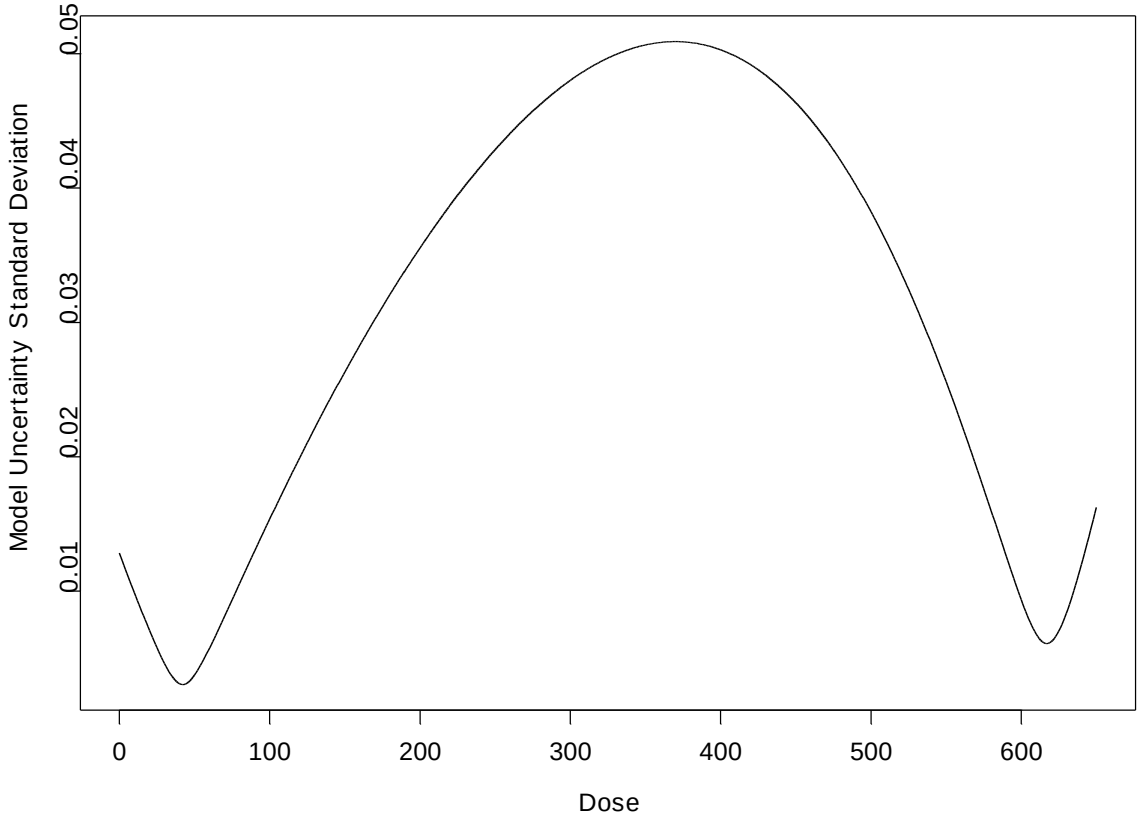


Figure 4. Plot of estimated probability from the best fitting model (M3 – only contains a linear dose term) and the BMA predicted probability based on fitting the 32 multistage models fit to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene [Data source: NTP (1983).]. The observed proportions of tumor-bearing female mice are plotted as "+".

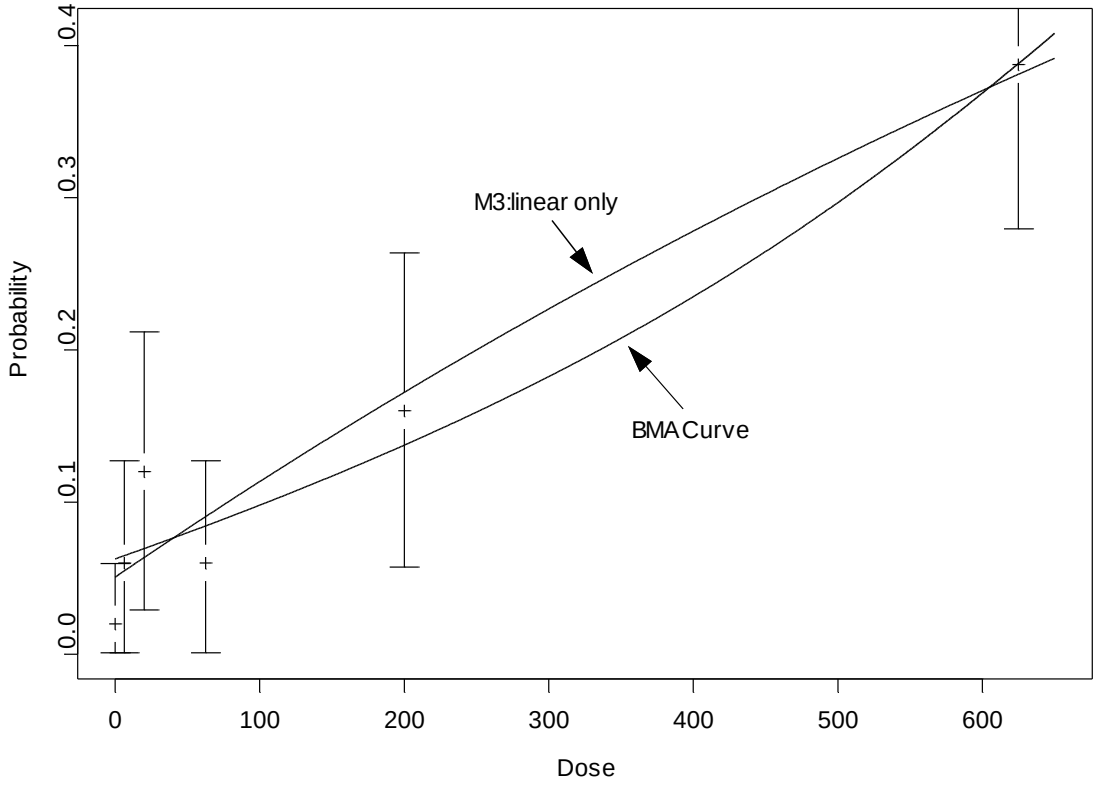
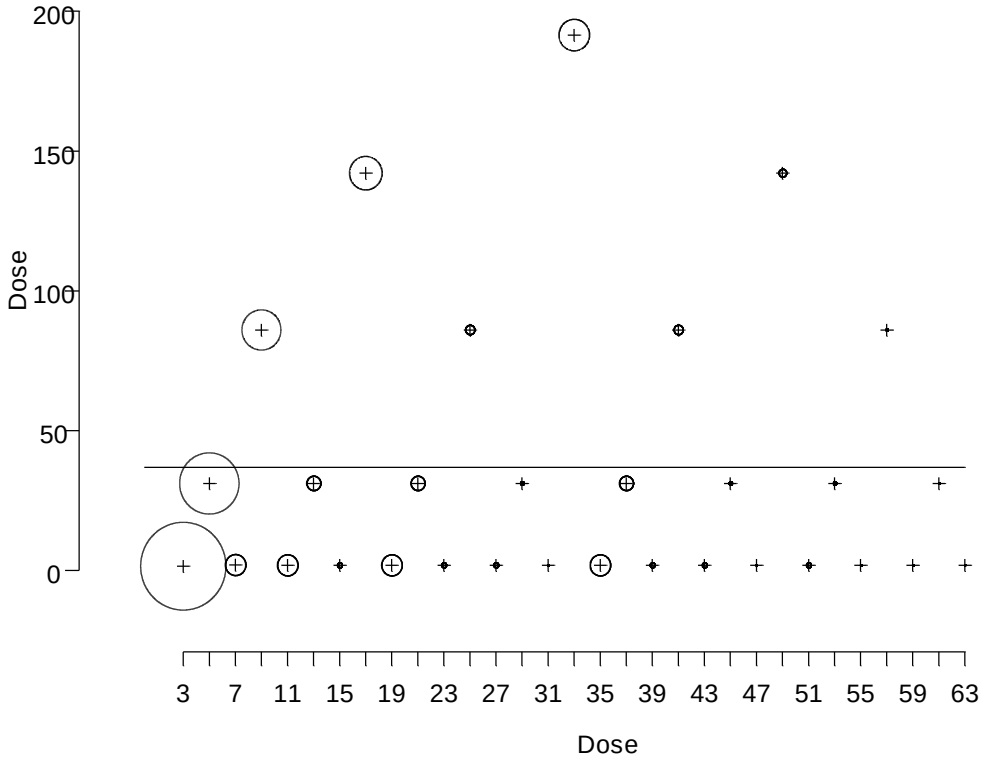


Figure 5. Plot of the estimated dose associated with an excess risk of 1 in a 1000 for each of the different models fit to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene [Data source: NTP (1983).]. The radius of each plotted circle is proportional to the posterior probability of the model. A horizontal reference line is plotted at the model average estimated dose associated with the excess risk of 0.001 –MA-BMD = 36.9.



Example 2 – Benchmark dose estimation with different models

As before, **BMD** = dose associated with a specified increase in response relative to the control response (**BMR**).

BMD proposed as response to NOEL/NOAEL

For example, if BMR=10%, then the BMD is the dose associated with a 10% increase in response relative to the controls.

BMDL = lower confidence limit on this dose

USEPA software provides BMD/BMDL estimates for dichotomous responses using a range of models including (1.2-1.10) below. Other models are available for continuous responses.

USEPA-BMD software models for binary response include:

log-logistic models: $\pi_i = \gamma + \frac{(1-\gamma)}{1+\exp\left[-\left(\alpha+\beta\ln(d_i)\right)\right]}$ (1.2)

gamma: $\pi_i = \gamma + (1-\gamma)\frac{1}{\Gamma(\alpha)}\int_0^{\beta d_i} t^{\alpha-1}e^{-t}dt$ (1.3)

multistage (degree=1) $\pi_i = \gamma + (1-\gamma)[1-\exp(-\theta_1 d_i)]$ (1.4)

multistage (degree=2) $\pi_i = \gamma + (1-\gamma)\left[1-\exp(-\theta_1 d_i - \theta_2 d_i^2)\right]$ (1.5)

probit $\pi_i = \Phi(a+\beta d_i)$ (1.6)

log-probit $$\pi_i = \gamma + (1-\gamma)\Phi[a + \beta \ln d_i] \quad (1.7)$$

quantal-linear $$\pi_i = \gamma + (1-\gamma)\left[1 - \exp(-\beta d_i)\right] \quad (1.8)$$

quantal-quadratic $$\pi_i = \gamma + (1-\gamma)\left[1 - \exp(-\beta d_i^2)\right] \quad (1.9)$$

Weibull $$\pi_i = \gamma + (1-\gamma)\left[1 - \exp(-\beta d_i^{\alpha})\right] \quad (1.10)$$

where $\Gamma(\alpha)$ = gamma function evaluated at α, $\Phi(x)$ = CDF for N(0,1) and $\pi_i = \gamma$ when $d_i=0$ for models (1.2) and (1.7).

Note: The multistage models (1.4) and (1.5) are equivalent to the previous form used on slide 10.

Example 2 – data illustration

Suppose that we want to compare the benchmark dose for a collection of different models

Data on renal tubular degeneration in rats exposed to Ethylene glycol (EG)

Responses: 2/10, 2/10, 2/10, 6/10, 9/10 for 0, 0.5, 1.0, 2.0 and 4.0% EG, respectively.

Risk endpoint: BMD and BMDL for BMR=1% and 10%.

Table. Summary of fitting different quantal response models to proportions of rats exhibiting renal tubular degeneration following exposure to ethylene glycol [Data source: Robinson et al. (1990).] (n=50 for all models). Columns in this table include BMD and BMDL for 10% and 1% BMR, a goodness of fit summary (P-value Pearson X^2), the Bayesian Information criterion (BIC), and a posterior model probability [$pr(M_n|D)$].

Model (M_m)	**BMD 10%**	**BMDL 10%**	**BMD 1%**	**BMDL 1%**	**P-value Pearson X^2**	**-2Log(L)**	**BIC**	***pr(M_n|D)***
Logistic	0.66	0.48	0.08	0.06	0.81	50.96	58.79	0.1736
Log-Logistic	1.15	0.43	0.55	0.09	0.89	50.23	61.97	0.0354
Gamma	1.09	0.32	0.49	0.34	0.82	50.40	62.13	0.0326
Multi Stage Linear	0.34	0.21	0.03	0.02	0.39	53.20	61.02	0.0567
Multi Stage Quadratic	0.95	0.28	0.29	0.03	0.90	50.61	58.44	0.2070
Probit	0.63	0.48	0.07	0.04	0.79	51.05	58.88	0.1660
Log-Probit	1.17	0.46	0.67	0.16	0.89	50.21	61.95	0.0357
Quantal-Linear	0.34	0.20	0.03	0.02	0.39	53.20	61.03	0.0566
Quantal-Quadratic	0.95	0.70	0.29	0.22	0.90	50.61	58.44	0.2070
Weibull	0.96	0.30	0.30	0.03	0.74	50.61	62.35	0.0293
BMA	**0.80**	**0.44**	**0.22**	**0.09**				

Observations:

* all models resulted in a *p*-value > 0.39 for a Pearson X^2 goodness-of-fit test. [conversations at NRC SEG]

* BMD10 ranged from 0.34 to 1.17% EG / BMDL10 from 0.20 to 0.70% EG, a four-fold range in BMDL.

* BMD01 ranged from 0.03 to 0.67% EG / BMDL01 from 0.02 to 0.34% EG, a 17-fold range in BMDL.

* estimated BMD or BMDL was more sensitive to model form for BMR = 1% as compared to BMR = 10%

* issue: but does it work? Stay tuned.

Example 3 - Variable selection and dose form?

Coal dust exposure in cross-sectional survey of miners

Response variable:

- respiratory lung function measured by the forced expiratory volume in one second (FEV_1).

Predictor variables:

- age of the miners (years)
- indicator of smoking (1 = smoker, 0 = otherwise)
- pack years of smoking
- indicator for race (1 = white, 0 = other)
- height of the miners (inches)
- indicator for western mines (1 = western, 0 = other)
- cumulative exposure to coal dust (mg-yr/m^3)

FEV_1 modeled as ~Normal with a linear model:

$$\hat{\mu}_e = b_0 + \sum_{i=1}^{k} b_k x_k + b_e f(e)$$

BMC definitions?

BMR = $\hat{\mu}_0 - \hat{\mu}_e$

BMC(AEC) = $f^{-1}\left[\frac{\delta}{-b_e} + f(0)\right]$

BMR = $\theta = \frac{\hat{\mu}_0 - \hat{\mu}_e}{\hat{\mu}_0}$

BMC(REC) = $f^{-1}\left[\frac{\theta \mu_0}{-b_e} + f(0)\right]$

Exposure form?

f(*e*): *e* (identity), $\sqrt{e}$ (square root), e^2 (square), and ln(*e*+ 1).

Model space?

* all possible combinations of the remaining predictors along with all possible two-way interactions of the predictors with the exposure metric while maintaining the hierarchy of terms (i.e., if an interaction is included in a model, then each of the terms in the interaction is included).

$$\mu_e = \beta_0 + \beta_1 age + \beta_2 height + \beta_3 \sqrt{e} + \beta_4 age \sqrt{e}$$

* 2916 models possible

Filtering models?

Occam's window (Madigan and Raftery) where models retained are in the set:

$$A = \left\{ M_k : \frac{max\{P(M_{ki}|D)\}}{P(M_k|D} \leq c \right\}$$

Used c= 20, and we opted to use the same constant. Thus, every model with posterior model probability less than $(1/20)^{th}$ the probability of the best model was excluded from the averaging.

Occam's razor:

- eliminate any model less likely than a model it is nested within.

Model-averaged BMC:
weighted average of the model-specific BMCs with weights reflecting the posterior model probabilities

$$BMC_{MA} = \sum \widehat{BMC}_i \times P(M_i|D)$$

Table I. Models That Survived the Occam's Window and Occam's Razor Filters Where "X" Indicates Term Is in the Model

Model	Expos Metric[a]	Ht.	Age	Smoking	Pack-Year	Race	Western	Interaction	R^2	BIC	P(M | D)
1	Identity	X	X	X	X	X	X	Height * Exposure	0.529	2537.6	0.39
2	Identity	X	X	X	X	X	X		0.527	2538.2	0.29
3	Square	X	X	X	X	X	X		0.526	2538.6	0.23
4	Square	X	X	X	X	X	X	Height * Exposure	0.528	2538.8	
5	Square root	X	X	X	X	X	X		0.527	2540.5	0.09
6	Square root	X	X	X	X	X	X	Height * Exposure	0.528	2541.1	

[a]Based on cumulative dust exposure.

Table II. BMC (Absolute Exposure) (mg/m^3 dust) Where the Model Numbers Below Refer to Models Described in Table I

Decrease in FEV_1	Height (in)	Model 1	Model 2	Model 3	Model 5	BMA
50 mL	68	60.1	44.2	93.5	11.4	58.8
		(29.7, 90.5)	(28.2, 60.3)	(76.2, 110.8)	(1.8, 21.1)	(7.8, 109.7)
	69	44.5	44.2	93.5	11.4	52.8
		(28.4, 60.7)	(28.2, 60.3)	(76.2, 110.8)	(1.8, 21.1)	(2.8, 102.8)
	71	29.3	44.2	93.5	11.4	46.9
		(21.5, 37.1)	(28.2, 60.3)	(76.2, 110.8)	(1.8, 21.1)	(0, 101.7)
100 mL	68	120.2	88.5	132.2	45.8	107.0
		(59.4, 181)	(56.5, 120.5)	(107.7, 156.7)	(7.3, 84.3)	(39.9, 174.1)
	69	89.0	88.5	132.2	45.8	94.9
		(56.7, 121.3)	(56.5, 120.5)	(107.7, 156.7)	(7.3, 84.3)	(38.8, 151)
	71	58.6	88.5	132.2	45.8	83.2
		(43, 74.2)	(56.5, 120.5)	(107.7, 156.7)	(7.3, 84.3)	(18.3, 148.1)

Notes: Table entries are BMC estimates along with 95% confidence intervals. The last column contains the model-averaged BMC point estimate and CI.

4. Does it work?

So, how well does MA work relative to standard practice?

Monte Carlo simulation study was conducted.

* "average-model" MA method contrasted with "average-dose" MA method of Kang *et al.* and Bailer *et al.* early work
* Design points:
  (0.0, 0.25, 0.5, 1.0) and (0.0, 0.0625, 0.125, 0.25, 0.5, 1.0).
* 50 "animals"/experimental units assigned to each dose group
* Given true D-R curve $\pi_{TRUE}(d)$, experiment was simulated and dichotomous adverse response generated and recorded.
* Considered 6 dose-response patterns covering 2 background response conditions crossed with 3 D-R steepness conditions (see Table III).

Table III. Six Response Patterns for the Doses 0, 0.5, and 1.0, Respectively, Which Were Used to Generate the True Underlying Curvature for the Simulation

Response Pattern	Response Curvature
1	(2%, 7%, 20%)
2	(2%, 14%, 34%)
3	(2%, 25%, 50%)
4	(7%, 12%, 25%)
5	(7%, 21%, 41%)
6	(7%, 32%, 55%)

* patterns allowed for curvature that ranged from sublinear to linear dose-response patterns.

* 54 diff. conditions=6 response patterns x 9 D-R models.

* responses ~ $Bin[n= 50, \pi_{TRUE}(d)\ ]$

* simulation = a set of dose-response curves being fit to the data with the corresponding MA-BMD as well as the MA-BMDL being estimated and recorded.

* 2000 random experiments generated through the simulation and bootstrap resampling (B=2000)

* Note:

- BMD couldn't be estimated for some data sets/bootstrap resamples
- Case where not possible to estimate a BMD, i.e., the BMD estimate is infinite due to a flat dose-response curve.
- While such a data set may not be appropriate for risk estimation, it cannot be dismissed in a bootstrap. These values are assigned an arbitrarily large value and we look at the estimates using the median.

* Summary values:

Coverage: Pr $(BMDL_{MA} \leq BMD_{TRUE})$

Obs. absolute relative median bias (i.e., median$\left\{\frac{\widehat{BMD}-BMD}{BMD}\right\}$)

Models fit:

Model averaged dose-response model:

$$\hat{\pi}_{MA}(d) = \sum_{k=1}^{K} \pi_k\left(\hat{\theta}_k, d\right) \times w_k$$

where

$$w_k = \frac{\exp\left(-\frac{1}{2} I_k\right)}{\sum_{i=1}^{K} \exp\left(-\frac{1}{2} I_i\right)}$$

where I_i represents the penalized information criterion described above (e.g., AIC or BIC).

- MA-BMD derived from numerical root-finding methods (method of bracketing and bisection)

- MA-BMDL from parametric bootstrap

Contrasts with an "average-dose" MA for BMD estimation, i.e.

$$\widehat{BMD}_{MA} = \sum_{k=1}^{K} BMD_k \times w_k$$

and

$$\widehat{BMDL}_{MA} = \sum_{k=1}^{K} BMDL_k \times w_k$$

Table II. Titanium Dioxide (TiO_2) Particle Surface Area Lung Burden Benchmark Dose (BMD/BMDL) Estimates, in Meters Squared per Gram of Lung (m^2/g), Which Corresponds to Extra Risks of 10% and 1% for Nine Popular Models (Models (2.1)–(2.9)). The Last Two Lines Correspond to the Model Averaged BMD Using the Averaged Dose-Response and Average-Risk Methods, as Well as Their Estimated Lower Bounds

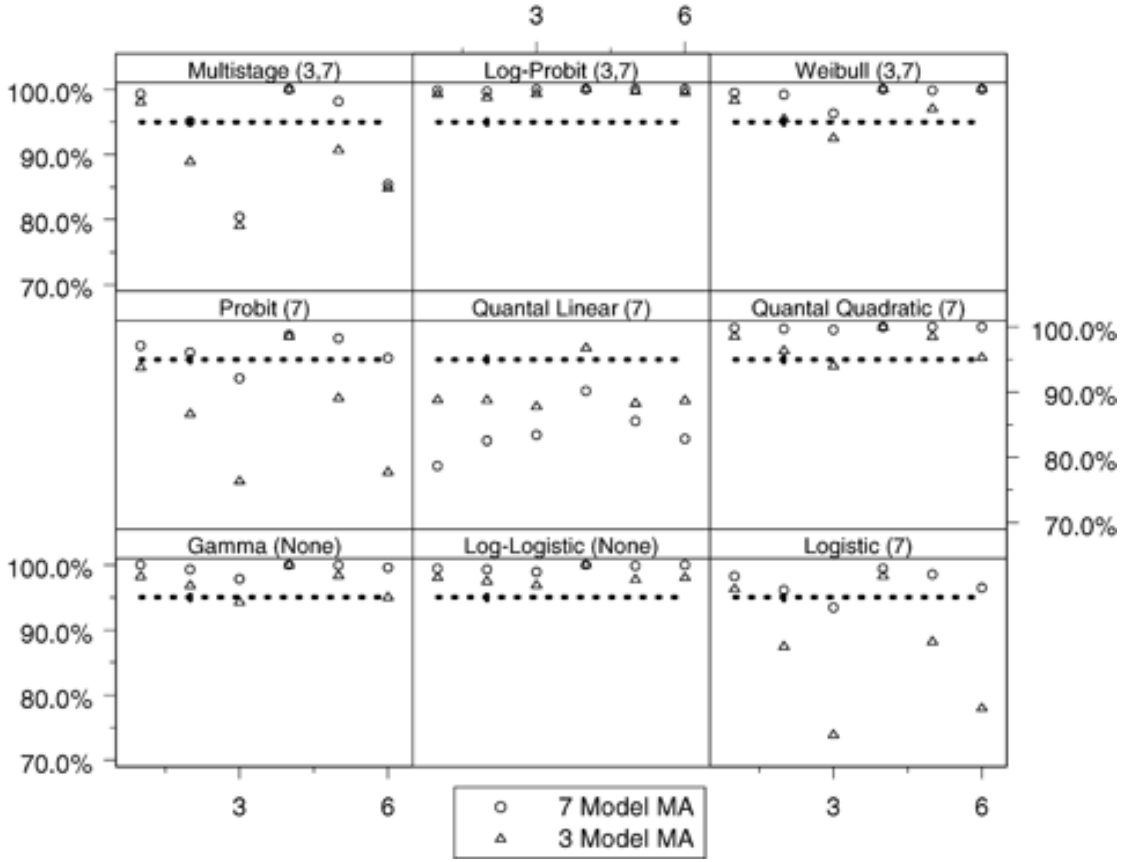
				Extra Risk BMD (BMDL)	
	Weights	**AIC**	**χ^2 GOF *p*-value**	**10%**	**1%**
Quantal-quadratic	0.192	364.95	0.57	0.96	0.30
				(0.84)	(0.26)
Logistic	0.154	365.39	0.50	1.01	0.25
				(0.92)	(0.2)
Probit	0.134	365.66	0.48	0.97	0.22
				(0.87)	(0.18)
Log-probit	0.129	365.74	0.55	1.01	0.51
				(0.79)	(0.20)
Gamma	0.119	365.90	0.53	1.03	0.52
				(0.83)	(0.18)
Log-logistic	0.112	366.03	0.52	1.04	0.50
				(0.82)	(0.18)
Weibull	0.108	366.11	0.52	1.04	0.49
				(0.83)	(0.17)
Multistage	0.039	368.11	0.61	1.04	0.47
				(0.85)	(0.14)
Quantal-linear	0.013	370.35	0.26	0.80	0.08
				(0.62)	(0.06)
"Average-model" model averaging BMD (BMDL)	n/a	n/a	n/a	1.01 (0.82)	0.36 (0.12)
"Average-dose" model averaging BMD (BMDL)	n/a	n/a	n/a	1.00 (0.84)	0.38 (0.20)

Two different model spaces:

3 model family = the multistage (2.1.4), the Weibull (2.1.9), and the log-probit (2.1.6)

7 model family = 3 model family + quantal-linear (2.1.7) + quantal-quadratic (2.1.8)+ logistic (2.1.1),+ probit (2.1.5)

Fig. 1. Observed coverage, that is, Pr(BMDL ≤ BMD) for BMR=1%. Within each panel, the horizontal axis corresponds to the six background response-steepness patterns given in Table III. The expected coverage, which is 95% for all cases, is represented by the dotted line and the observed coverage is represented as points in the plot. The numbers in parentheses represent the model spaces when the TRUE model was included in the simulation (i.e., the three- or seven-model space); none implies the model was not included in any model space.

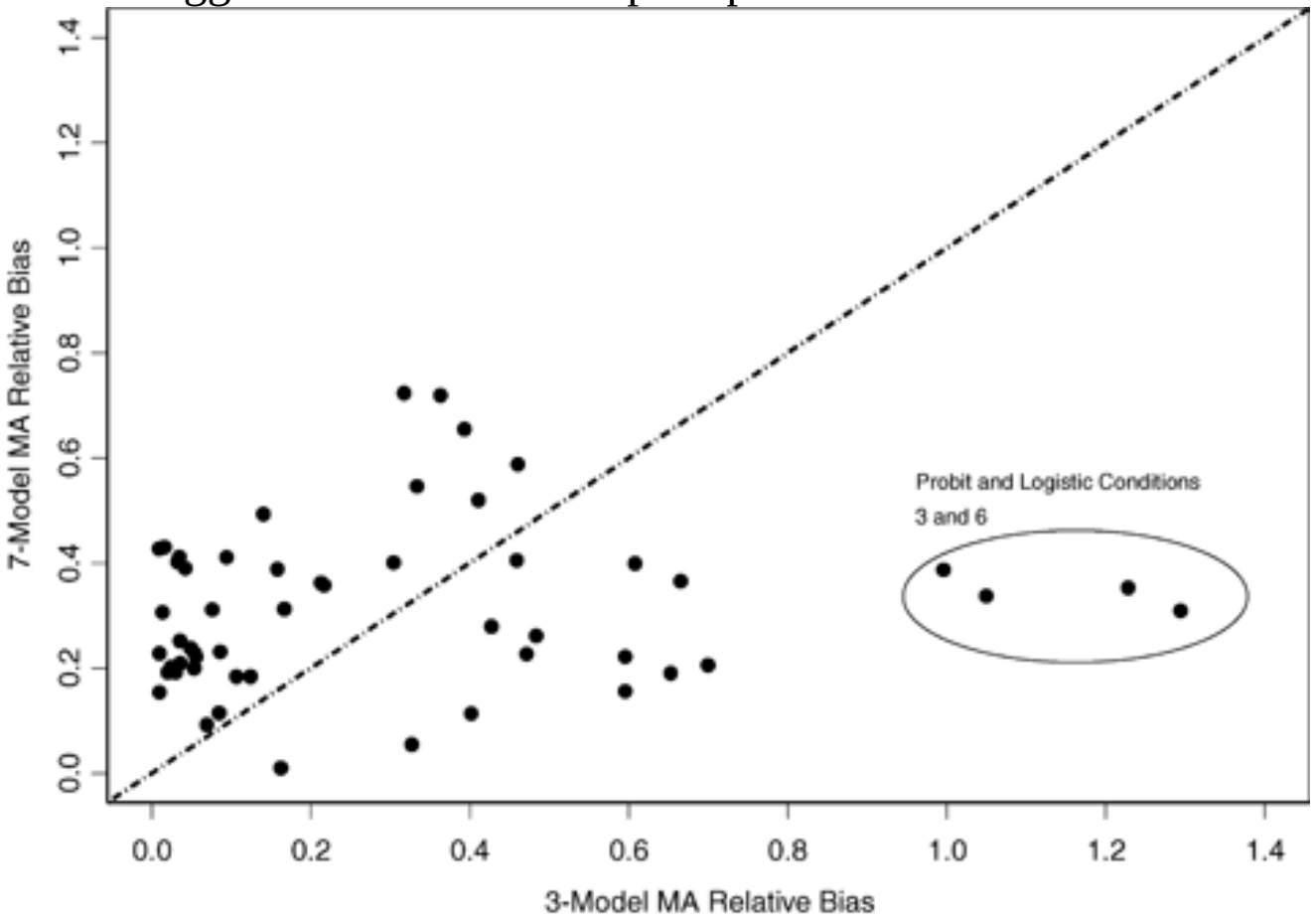


Problem: MA with a true quantal-linear model (also evident in the model parameterizations that have near linear dose-response patterns)

* closer the simulation condition was to a linear dose-response the poorer the MA performance.

-> See multistage model where the third and sixth simulation patterns are nearly linear (i.e., the linear term dominates the quadratic term in the range of the data).

Fig. 3. Comparison of the absolute relative bias between the BMD estimates computed using the three- and seven-model spaces for BMRs of 1%. Points above the line suggest the seven-model space produced more biased results and points below suggest the three-model space produced more biased results.



Wheeler and Bailer[6] also compared looked at the coverage behavior of

1) BMD based on picking the "best model", i.e., the model with the largest *p*-value is picked as the "best model."

vs.

2) BMD based on “average model” MA estimate

@BMR=10%:

* MA performs at or above the nominally specified 95% level in 34 of the 54 simulations for the three-model conditions
* Best model 9 of the 54 conditions.
* “average-dose” MA estimate reached nominal level in 15 of 54 conditions

So, what's with the quantal linear model?

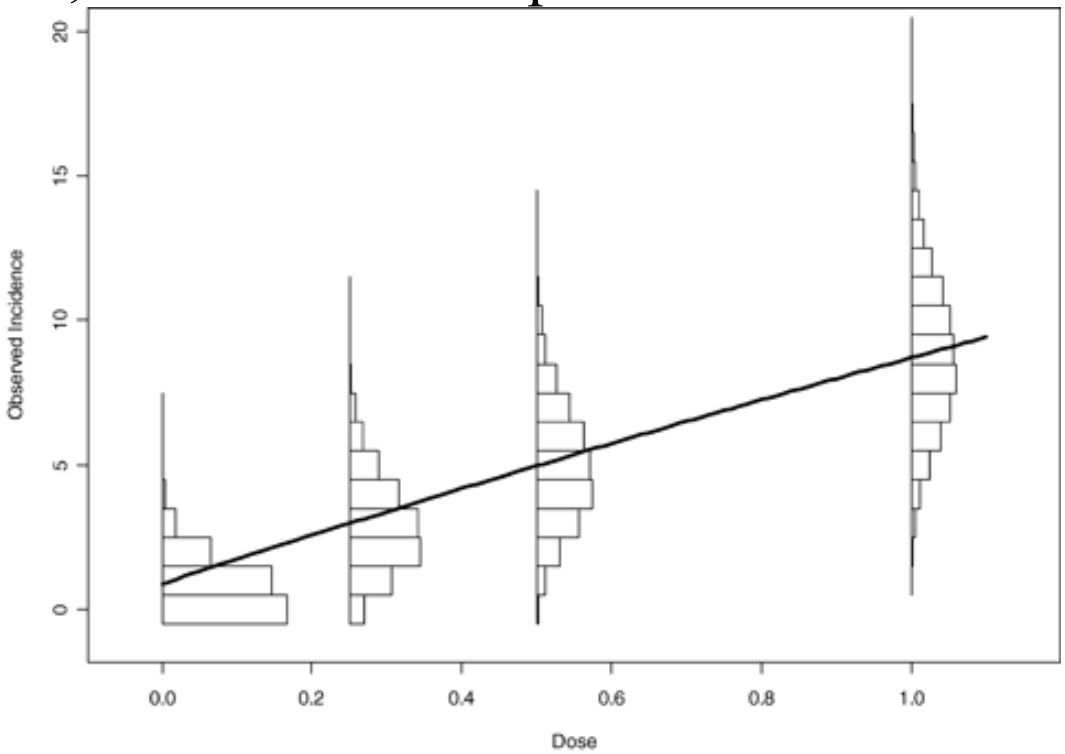


Fig. 5. Sampling distribution for each dose group in the simulation design, with the expected quantal-linear response superimposed. Each histogram describes the distribution of observed response where 50 animals are exposed to doses of 0, 0.25, 0.5, and 1.0 with the probability of a positive response is 0.017, 0.059, 0.10, and 0.17, respectively. The skewed distributions at the lower doses may help to explain the coverage and bias behavior of model averaging in the quantal-linear case.

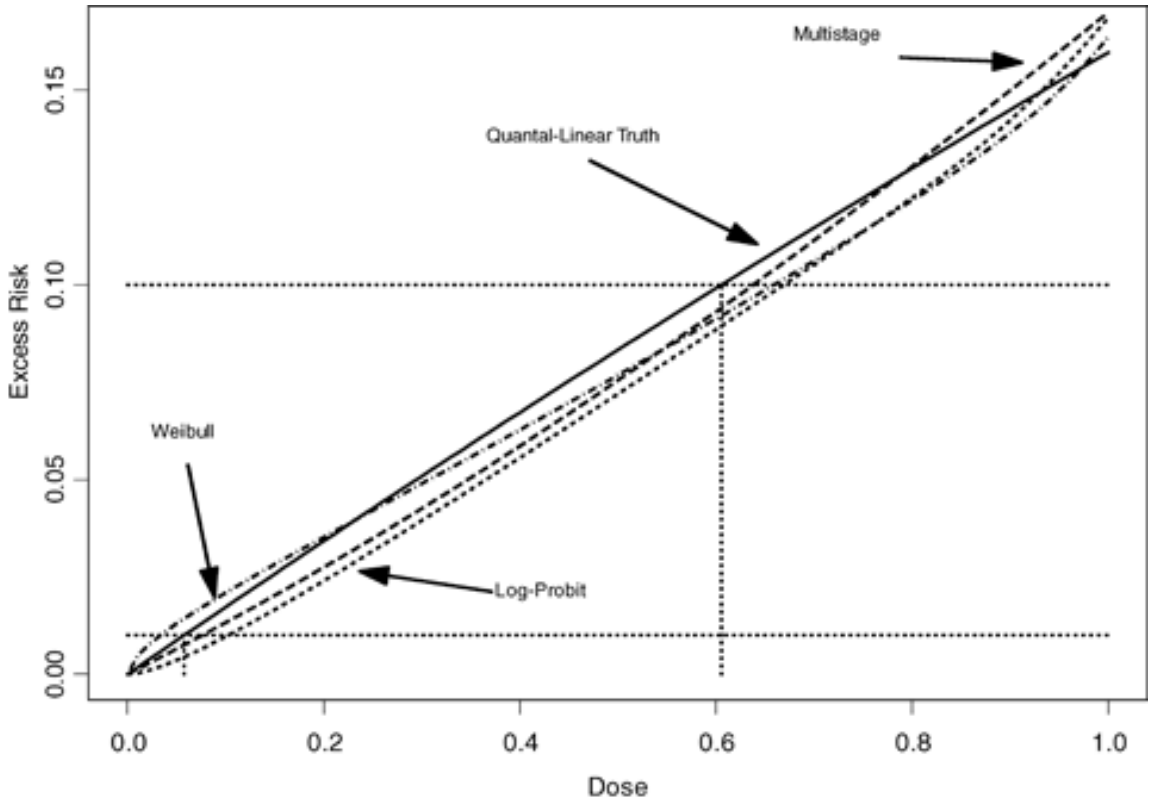


Fig. 6. The average excess risk curve of 2,000 model fits for the Weibull, multistage, and the log-probit models. The models were fit to 2,000 data sets generated from the first quantal-linear condition. These average curves show a consistent underrepresentation of the true risk, and corresponding benchmark dose for the three models involved. The horizontal lines correspond to BMRs of 1% and 10%, while the vertical lines correspond to the BMD estimate.

4. Future [Matt's recent work]

*These methods all assume an averaged parametric form for the underlying dose-response model.

*We are investigating methods that are fully nonparametric in the dose response form [spline model with a monotonicity constraint]

*This is difficult due to the implied monotonicity, and limited data (particularly in the Quantal data case).

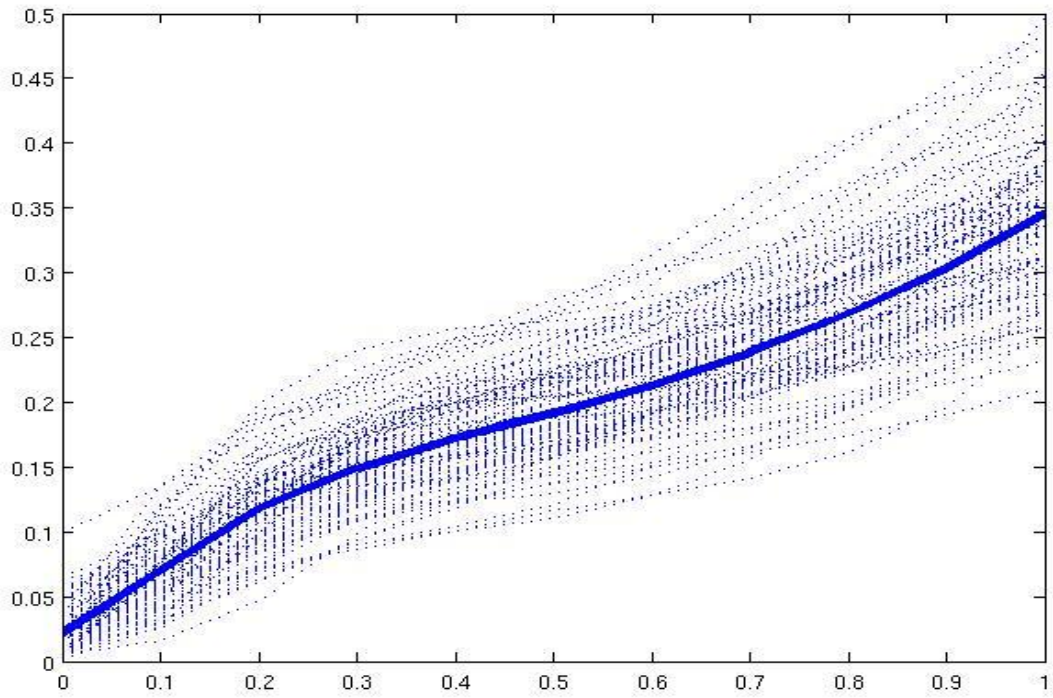


Fig 7. A fully nonparametric fit for a quantal data set that was derived from a linear model fit. The model assumes monotonic B-splines fit on a probit scale.

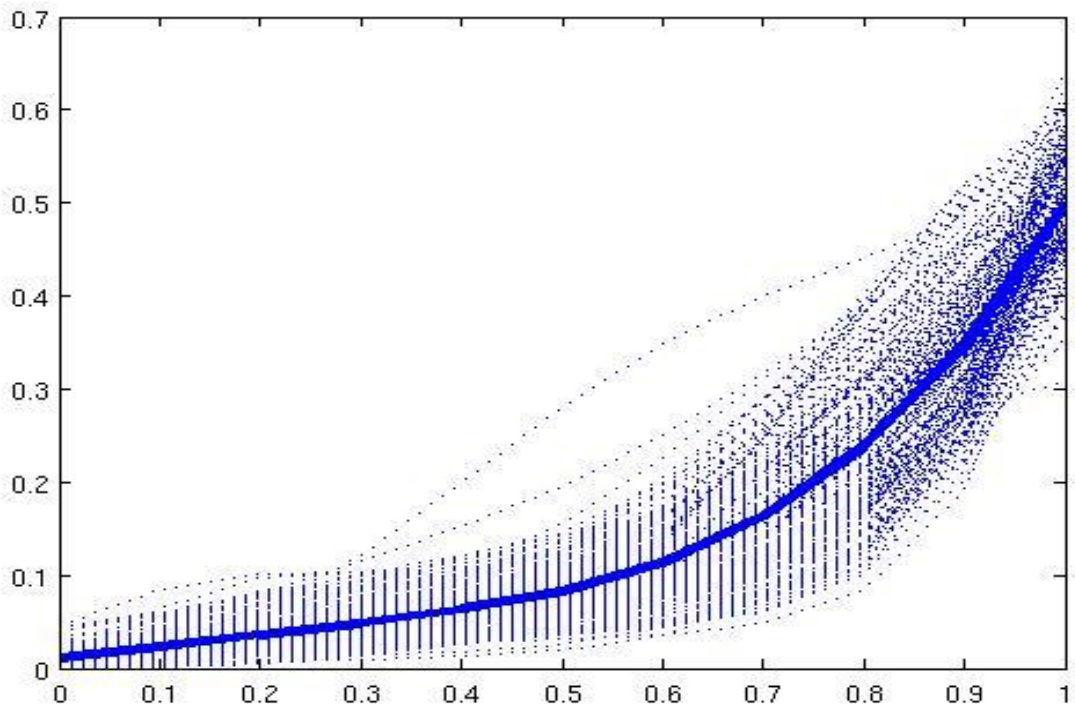


Fig 8. A fully nonparametric fit for a quantal data set that was derived from a sublinear model. The model is the same as above assuming only monotonic B-splines fit on a probit scale.

* These methods do show some promise.

* Preliminary results suggest nominal coverage is met for the quantal linear model.

* This is the first time we have observed nominal coverage when using a method that did not specify the underlying model as quantal-linear.

To recap …

Time	QRA and risk estimation based on …	Observations …
Past (current?) practice	Single model	Simple, Few parameters, Interpretable, Overly optimistic inference
↓	MA (dose averaging)	Simple to implement with existing software, Poor coverage
↓	MA (D-R model averaging)	Relatively simple although customized software needed; better operating characteristics
↓		
Future	Flexible D-R model forms (e.g. semiparametric smooth models)? Biologically-based models	Most flexibility; Model parameters? Subtle to implement; complex software

COMMENTS:

* Model uncertainty needs to be addressed in the risk estimation process.

* Model averaging is an attractive option for addressing this.

* Did not emphasize a full Bayesian analysis of the system where posterior probability distributions of the regression parameters were also of interest.

* One cost [benefit?] associated with MA is that now risk estimates from different models are explicitly considered.

* With this strategy, the risk modeler is not forced to select a particular model for generating risk estimates.

* Currently, many risk modelers report estimates based upon numerous model fits and then report a preference for a particular model and estimate, perhaps for the model leading to the lowest exposure limit.

* Model averaging approach, a collection of models is selected in advance and estimates derived from these models are weighted by the support the data suggest for each model.

* Dose-response models are selected prior to generating an excess risk estimate. Model averaging allows for a synthesis of risk estimates where the estimate from each model is weighted by a posterior probability of the model.

Using BMA in risk estimation?

* Report the risk endpoint for the most probable model along with the BMA estimate for the same endpoint.

* If these quantities are similar or if the model uncertainty variance component is small, then the risk assessor may have greater confidence that the reported risk endpoint is fairly robust to model uncertainty.

* If not, then the risk manager may wish to consider selecting a different endpoint (e.g. BMD10 vs. BMD01) or to provide other evidence supporting the selection of a particular model for risk estimation.

* We believe that the best way to incorporate BMA in risk assessment should continue to be debated.

* Ultimately, risk estimation that ignores the uncertainty associated with model selection may be present itself as overly precise, and we hope that the strategy discussed above may help address this.

* Does not necessarily simplify the risk estimation process, it does help account for the impact of model uncertainty on risk estimates in a natural way. As Kass and Raftery(1995) noted:

  Any approach that selects a single model and then makes inferences conditionally on that model ignores the uncertainty involved in model selection, which can be a big part of overall uncertainty. This leads to underestimation of the uncertainty about quantities of interest, sometimes to a dramatic extent. [Kass and Raftery,1995, p. 784]

Conclusions

* Box was quoted as saying “all models are wrong but some are useful”; my colleague Bob Noble has described BMA as process that says “why not use all useful models.”

* Addressing uncertainty and variability in risk assessment does not necessarily simplify the process of RA; however it may provide a better reflection of what is unknown in the process.

Partial list of references

- 1. Kang, S. H., Kodell, R. L., Chen, J. J. (2000). Incorporating model uncertainties along with data uncertainties in microbial risk assessment. *Regulatory Toxicology and Pharmacology* , ***32*** , 68–72. Links
- 2. Moon, H., Hyun-Joo, K., Chen, J. J., Kodell, R. L. (2005). Model averaging using the Kullback information criterion in estimating effective doses for microbial infection and illness. *Risk Analysis* , ***25*** , 1147–1159. Links
- 3. Bailer, A. J., Noble, R. B., Wheeler, M. (2005). Model uncertainty and risk estimation for quantal responses. *Risk Analysis* , ***25*** , 291–299. Links
- 4. Bailer, A. J., Wheeler, M., Dankovick, D., Noble, R., Bena, J. (2005). Incorporating uncertainty and variability in the assessment of occupational hazards. *Internation Journal of Risk Assessment and Management* , ***5*** , 344–357. Links
- 5. Morales, K. H., Ibrahim, J. G., Chen, C. J., Ryan, L. M. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *Journal of the American Statistical Association* , ***101*** , 9–17. Links
- 6. Wheeler, M. W., Bailer, A. J. (2009). Comparing model averaging with other model selection strategies for benchmark dose estimation. *Environmental and Ecological Statistics* , ***16*** (1). Links
- 7. U.S. Environmental Protection Agency. (2001). *Help Manual for Benchmark Dose Software Version:1.3* . EPA 600/R-00/014F. Research Triangle Park , NC : USEPA.
- 8. Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundemental and Applied Toxicology* , ***4*** , 854–871. Links
- 9. Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology* , ***25*** , 111–163. Links
- 10. Buckland, S. T., Burnham, K. P., Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* , ***53*** , 603–618. Links
- 11. Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics* , ***30*** , 9–14. Links
- 12. Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* , ***6*** , 461–464. Links
- 13. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing* , 2nd ed. Cambridge : Cambridge University Press.
- 14. Efron, B., Tibshirani, R. B. (1993). *An Introduction to the Bootstrap.* New York : Chapman & Hall.
- 15. Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association* , ***82*** , 171–185. Links