

Model Uncertainty and Risk Estimation for Experimental Studies of Quantal Responses

A. John Bailer,^{1,2*} Robert B. Noble,¹ and Matthew W. Wheeler²

Experimental animal studies often serve as the basis for predicting risk of adverse responses in humans exposed to occupational hazards. A statistical model is applied to exposure-response data and this fitted model may be used to obtain estimates of the exposure associated with a specified level of adverse response. Unfortunately, a number of different statistical models are candidates for fitting the data and may result in wide ranging estimates of risk. Bayesian model averaging (BMA) offers a strategy for addressing uncertainty in the selection of statistical models when generating risk estimates. This strategy is illustrated with two examples: applying the multistage model to cancer responses and a second example where different quantal models are fit to kidney lesion data. BMA provides excess risk estimates or benchmark dose estimates that reflects model uncertainty.

KEY WORDS: Bayesian model averaging; benchmark doses; quantal multistage models; unit cancer risk

1. INTRODUCTION

Animal studies are often used as the basis for estimating the risks associated with exposure to environmental and occupational hazards. Commonly, animals are randomly assigned to receive a particular level of a hazard for a specified period of time. An adverse response (e.g., tumor development, renal tubular degeneration) is recorded for each animal. For dichotomous responses, the probability of the adverse response is modeled as a function of the hazard concentration. Given a specified model, the additional risk or extra risk above background at a specified dose or the dose associated with a specified response (e.g., the benchmark dose or BMD) is estimated. This calculation will vary based on the selected

regression model. The availability of software such as the USEPA benchmark dose software⁽¹⁾ facilitates the fitting of a collection of dose-response models. After fitting these models and deriving associated extra risks or benchmark doses, the risk manager is faced with the question of which model should be used, or equivalently, how model uncertainty in the risk estimation process should be addressed. One option is to select the model that leads to the highest extra risk or the lowest benchmark dose with the belief that this selection is biased toward being protective. Another option is to report a range of risk estimates for models that provide a good fit to the observed data. This provides a sensitivity analysis of the extra risk/BMDs as a function of the regression models used. While selecting benchmark responses at levels in the 5–10% range tends to mitigate differences in the estimated BMDs, differences will still exist, especially in the lower limits on the BMD. We consider a third option for dealing with model-related uncertainty in risk estimates—averaging of risk estimates/BMDs based on the support for each model provided by the data. In the sections that follow, we outline Bayesian Model

¹ Department of Mathematics and Statistics, Miami University, Oxford, OH 45056, USA.

² Risk Evaluation Branch, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, OH 45224, USA.

* Address correspondence to A. John Bailer, Department of Mathematics and Statistics, Miami University, Oxford, OH 45056, USA; tel: 513-529-3538; fax: 513-529-1493; baileraj@muohio.edu.

Averaging (BMA) as applied to quantal response models commonly employed in risk assessment.

2. METHODS

2.1. Notation

Suppose g nonzero concentration conditions (d_1, \dots, d_g) along with a control condition ($d_0 = 0$) are examined in a study of the carcinogenic effects of some potentially hazardous compound. The number of organisms exposed to the i th condition is n_i . The number of organisms exhibiting the response of interest is recorded, say Y_i for the i th concentration condition. Finally, let π_i represent the population parameter corresponding to the true probability of developing the response of interest given exposure to hazard concentration d_i . In the examples that follow, we consider Y = number of animals with tumors or Y = number of animals with renal tubular degeneration.

2.2. Quantal Response Models

A host of models are available for modeling the probability of response (π) as a function of dose (d). For example, a common model for the probability of tumor presence as a function of carcinogen dose is the quantal multistage response model,⁽²⁾

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

where $k \leq g$ and the regression coefficients are frequently constrained to be nonnegative, i.e., $\theta_i \geq 0$. The number of terms in the model is usually less than the total number of nonzero concentration conditions. Other examples of quantal response models include the collection of regression models available in the benchmark dose software available from the USEPA. A subset of these models is given below, and these include:

log-logistic models:

$$\pi_i = \gamma + \frac{(1 - \gamma)}{1 + \exp[-(\alpha + \beta \ln(d_i))]} \quad (2)$$

gamma:
$$\pi_i = \gamma + (1 - \gamma) \frac{1}{\Gamma(\alpha)} \int_0^{\beta d_i} t^{\alpha-1} e^{-t} dt \quad (3)$$

multistage (degree = 1):

$$\pi_i = \gamma + (1 - \gamma)[1 - \exp(-\theta_1 d_i)] \quad (4)$$

multistage (degree = 2):

$$\pi_i = \gamma + (1 - \gamma)[1 - \exp(-\theta_1 d_i - \theta_2 d_i^2)] \quad (5)$$

probit:
$$\pi_i = \Phi(a + \beta d_i) \quad (6)$$

log-probit:
$$\pi_i = \gamma + (1 - \gamma)\Phi[a + \beta \ln d_i] \quad (7)$$

quantal-linear:
$$\pi_i = \gamma + (1 - \gamma)[1 - \exp(-\beta d_i)] \quad (8)$$

quantal-quadratic:

$$\pi_i = \gamma + (1 - \gamma)[1 - \exp(-\beta d_i^2)] \quad (9)$$

Weibull:
$$\pi_i = \gamma + (1 - \gamma)[1 - \exp(-\beta d_i^\alpha)], \quad (10)$$

where $\Gamma(\alpha)$ = gamma function evaluated at α , $\Phi(x)$ is the cumulative distribution function of the standard normal density at x (i.e., the integral of a $N(0, 1)$ density from $-\infty$ to x), and $\pi_i = \gamma$ when $d_i = 0$ for Models (2) and (7). Note that Models (4) and (8) are identical and either can be selected as options in the USEPA BMD software.

2.2.1. Extra Risk

A characterization of the extra risk of tumor development at a given dose can be defined as

$$ER(d_i) = \frac{\pi_i - \pi_0}{1 - \pi_0}. \quad (11)$$

In multistage Model (1) with a linear term plus other higher-order terms, this can be approximated as $ER(d) \approx \theta_1 d$ for doses close to 0. The parameter θ_1 is the unit cancer risk that corresponds to the estimated excess risk of tumor development associated with a unit increase in exposure. An alternative characterization of the impact of a chemical is the added risk, $AR(d_i) = \pi_i - \pi_0$.

The presence of a linear term in the model is clearly associated with higher risk of tumor development at lower exposures than a model without a linear term. The estimation of excess risk is contingent on the selection of a particular model, i.e., the terms included in Model (1). This excess risk calculation ignores the uncertainty inherent in the selection of this model. A strategy for incorporating model uncertainty in the estimation of excess risk is desirable as is the ability to

characterize the probability that a linear term would be included in risk estimation models.

2.2.2. Benchmark Doses

An alternative to extra risk is the benchmark dose.⁽³⁾ The BMD is the dose associated with a specified increase in response relative to the control response, the so-called benchmark response (BMR). For example, if $BMR = 10\%$, then the BMD is the dose associated with a 10% increase in response relative to the controls. A lower confidence limit on this dose is often calculated. The lower limit is called a benchmark dose lower limit (BMDL). The USEPA software provides BMD/BMDL estimates for dichotomous responses using a range of models, including Models (2)–(10). Other models are available for continuous responses.

2.3. Family of Models

Bayesian Model Averaging formally incorporates model uncertainty in an analysis. In BMA, a family of models is defined, say M , data are collected, and posterior probabilities associated with the models are derived. As an illustration, consider a quantal multistage Model (1) being fit to data containing $g = 6$ experimental groups. Multistage models (1) with all dose terms of order five or lower are possible. This collection of possible models is given in Table I.

Extra risk⁽²⁾ is zero for model ID = 1 since no dose terms are present in this model. Extra risk for model ID = 3 is $\Delta(d_i) = 1 - \exp(-\theta_1 d_i)$ while extra risk for model ID = 33 is $\Delta(d_i) = 1 - \exp(-\theta_5 d_i^5)$. Generally, extra risk for model ID = 3 will exceed extra risk for model ID = 33 for small dose levels. This occurs because model ID = 3 has a linear term while model ID = 33 has a fifth-order term as its lowest polynomial term. In the model with greater curvature (ID = 33), the slope of the dose response at doses near 0 will be much smaller than the slope of the model with the linear term (ID = 3). This translates into higher extra risk estimates for a particular low dose for models with linear dose terms. We could restrict the models that we consider as candidates for extra risk calculations. For example, if we enforce a hierarchy in which the presence of a higher-order term forces the inclusion of all lower-order terms, then only six plausible models for the probability of response remain (namely, Models 1, 3, 7, 15, 31, and 63 in Table I). Alternatively, we could restrict attention to models that contain a linear term and some combination of higher-order terms

(Models 3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, and 63 in Table I).

2.4. Bayesian Model Averaging

The basic idea in BMA is that the distribution of some characteristic of a model, e.g., excess risk or BMD/BMDL, is derived over some space of possible models.⁽⁴⁻⁶⁾ The posterior distribution of this characteristic, Δ , given the data are $\text{pr}(\Delta | D) = \sum_{m=1}^{N_m} \text{pr}(\Delta | M_m, D) \text{pr}(M_m | D)$, where the posterior probability of model M_m given the data D is $\text{pr}(M_m | D)$ and the number of possible models considered is N_m .

Two issues related to prior probability distribution need to be considered. First, what should be specified as the prior probability of each model, $\text{pr}(M_m)$? A common choice is to assume that all models are equally likely *a priori*. We believe this makes sense for most risk estimate applications of BMA. For example, with USEPA BMD software, users may view each option in the menu of plausible models as equally likely. The second issue is a little more subtle. Each model includes a set of parameters that need to be estimated. For example, Model (1) includes parameters $(\theta_0, \theta_1, \dots, \theta_k)$ while Model (10) includes (γ, α, β) . What prior distribution is placed on the parameters of these models? The marginal likelihood, $\text{pr}(M_m | D)$, is computed by integrating the likelihood multiplied by the prior distribution of the parameters over the parameter space. The computation of the integrals is generally difficult and as a result is often approximated using the BIC. The BIC provides a reasonable approximation of the marginal likelihood when the unit-information prior is assumed on the parameter space (see Hoeting *et al.*⁽⁶⁾ and Kass and Raftery⁽⁷⁾). This likelihood is needed to construct Bayes factors, the odds of preferring one model over another. Kass and Raftery⁽⁷⁾ suggest that the introduction of prior densities on the model parameters can be avoided by the use of the Schwarz criterion, which is a rough approximation to the log of the Bayes factor. The Bayesian Information Criterion (BIC) is a function of the log-likelihood (LL) of a model, number of observations (n), and the number of parameters in the model (k)—namely, $\text{BIC} = -2 \times \text{LL} + k \times \ln(n)$. (The Schwarz criterion S in Kass and Raftery⁽⁷⁾ is related to the BIC of two models, $S = -0.5 \times (\text{BIC}_1 - \text{BIC}_2)$ where BIC_i is the BIC associated with model $i = 1, 2$.) Schwarz⁽⁸⁾ derived the BIC as a large sample approximation to twice the logarithm of the Bayes factor. Kass and Wasserman⁽⁹⁾ showed that BIC

Table I. Summary of Fitting Different Multistage Models to Lymphocytic Malignant Lymphomas in Female Mice Exposed to 1,3-Butadiene (Data Source: NTP (1983)) ($n = 330$ for All Models)

Model ID (M_m) ^a	d	d^2	d^3	d^4	d^5	BIC	$\text{pr}(M_n D)$	p -Value	Pearson χ^2	Dose
1	0	0	0	0	0	70.561	0.00000	0.000		NA
3	1	0	0	0	0	35.656	0.42810	0.366		1.46
5	0	1	0	0	0	37.101	0.20790	0.264		30.00
7	1	1	0	0	0	41.377	0.02451	0.243		1.81
9	0	0	1	0	0	38.782	0.08973	0.143		83.64
11	1	0	1	0	0	41.364	0.02467	0.245		1.73
13	0	1	1	0	0	42.900	0.01144	0.264		30.00
15	1	1	1	0	0	47.163	0.00136	0.245		1.73
17	0	0	0	1	0	39.508	0.06241	0.107		139.10
19	1	0	0	1	0	41.364	0.02467	0.245		1.71
21	0	1	0	1	0	42.900	0.01144	0.264		30.00
23	1	1	0	1	0	47.163	0.00136	0.245		1.71
25	0	0	1	1	0	44.581	0.00494	0.143		83.64
27	1	0	1	1	0	47.163	0.00136	0.125		1.70
29	0	1	1	1	0	48.699	0.00063	0.264		30.00
31	1	1	1	1	0	52.962	0.00007	0.245		1.71
33	0	0	0	0	1	39.759	0.05505	0.096		188.19
35	1	0	0	0	1	41.364	0.02467	0.245		1.70
37	0	1	0	0	1	42.900	0.01144	0.264		30.00
39	1	1	0	0	1	47.163	0.00136	0.245		1.70
41	0	0	1	0	1	44.581	0.00494	0.143		83.64
43	1	0	1	0	1	47.163	0.00136	0.125		1.70
45	0	1	1	0	1	48.699	0.00063	0.264		30.00
47	1	1	1	0	1	52.962	0.00007	0.125		1.70
49	0	0	0	1	1	45.307	0.00344	0.107		139.10
51	1	0	0	1	1	47.163	0.00136	0.245		1.71
53	0	1	0	1	1	48.699	0.00063	0.264		30.00
55	1	1	0	1	1	52.962	0.00007	0.245		1.71
57	0	0	1	1	1	50.380	0.00027	0.143		83.64
59	1	0	1	1	1	52.962	0.00007	0.245		1.71
61	0	1	1	1	1	54.498	0.00003	0.264		30.00
63	1	1	1	1	1	58.761	0.00000	0.041		1.73

BMA = 36.01

^aModel ID number is based upon terms present in the model. In particular, the ID no. = $1 + 2 * "d^1" + 4 * "d^2" + 8 * "d^3" + 16 * "d^4" + 32 * "d^5"$, where " d^j " = 1 if dose term d^j is present in the model and " d^j " = 0 if not ($j = 1, 2, 3, 4,$ and 5).

Note: Columns in this table represent which terms are in the model (" d^i " = 1 if a dose term raised to the i th power is in the model and 0 otherwise), the Bayesian information criterion (BIC), posterior model probability [$\text{pr}(M_n | D)$], goodness-of-fit summary, and dose associated with an extra risk of 1 in a 1,000.

provides a close approximation to the Bayes factor when a unit-information prior on the parameter space is used. Raftery⁽¹⁰⁾ 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 $\exp(-0.5 \text{ BIC})$, i.e.,

$$\text{pr}(M_m | D) = \frac{\exp(-0.5 \text{ BIC}_m)}{\sum_{r=1}^{N_m} \exp(-0.5 \text{ BIC}_r)}$$

Once the posterior distribution of the characteristic given the data is derived, then any relevant

summary can be constructed. For example, the posterior mean dose estimate associated with an excess risk of 0.001 can be constructed as $E[\Delta(d) | D] = \sum_{m=1}^{N_m} \hat{\Delta}_m(d) \text{pr}(M_m | D)$, where $\hat{\Delta}_m(d)$ is the dose that yields $\text{ER}(d) = 0.001$ for model M_m .

2.5. Example 1: Lymphocytic Malignant Lymphomas in 1,3-Butadiene Exposed Mice

We illustrate the application of BMA methods to a data set describing lymphocytic malignant lymphomas observed in female mice exposed to 1,3-butadiene.⁽¹¹⁾ The observed quantal responses associated with exposures 0, 6.25, 20, 62.5, 200, and

625 ppm were 1/50 (2%), 3/50 (6%), 6/50 (12%), 3/50 (6%), 8/50 (16%), and 31/80 (39%), respectively (Table 6, NTP⁽¹¹⁾). Survival was also impacted by exposure to 1,3-butadiene with all exposure groups greater than 6.25 ppm exhibiting significantly decreased survival relative to controls; however, we are not considering this additional complexity in our example.

2.6. Example 2: Renal Tubular Degeneration in Ethylene Glycol Exposed Rats

A 10-day exposure study was conducted where Sprague-Dawley rats were exposed to ethylene glycol (EG) in their drinking water.⁽¹²⁾ Ten rats were exposed in one of five dose groups (0, 0.5, 1.0, 2.0, and 4.0% EG). The observed proportions of rats exhibiting renal tubular degeneration were 2/10, 2/10, 2/10, 6/10, and 9/10 for EG concentrations of 0%, 0.5%, 1.0%, 2.0%, and 4.0% EG, respectively.

3. RESULTS

3.1. Example 1: BMA for Excess Risk of Lymphocytic Malignant Lymphomas in 1,3-Butadiene Exposed Mice

In this section, we focus on the estimation of the dose associated with an excess risk of 1 in 1,000 accounting for model uncertainty. The selection of an excess risk of 1/1,000 is a common choice in the evaluation of occupational hazards. We consider that family of models to contain the 32 (=2⁵) described in Table I. A summary of these model fits is also included in Table I while a plot of these models fit to the female mouse lymphocytic malignant lymphoma data are given in Fig. 1. The fit of the intercept-only model (Model ID = 1 from Table I) is a horizontal line at approximately 0.15. This intercept-only model yields a model prediction that corresponds to the total number of cancers in all experimental groups (52) divided by the total number of animals studied (330). This model clearly does not provide a good characterization of the observed dose-quantal response data. From this figure, we see that many of the remaining 31 models do provide a good summary of this relationship.

These models have differing amounts of support from the data, as illustrated in Fig. 2. The posterior model probability for each of the models is displayed in this figure. The intercept-only model (Model

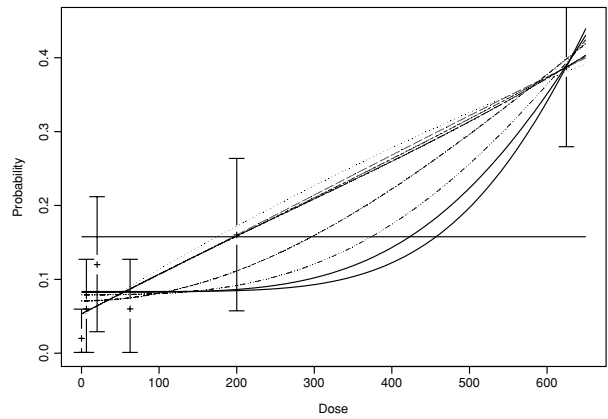


Fig. 1. Plot of the estimated probability of lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene (data source: NTP (1983)). Each curve is associated with the fit of each of the 32 candidate models. The observed quantal responses are plotted as “+”. Confidence intervals are plotted along with the observed proportions.

ID = 1) has a posterior model probability near 0, the model containing only a linear term (Model ID = 3) has the highest posterior model probability (approximately 0.43), and the quadratic only model (Model ID = 5) has a posterior model probability of 0.21. A number of these models have posterior model probabilities exceeding 0.05.

The standard deviation of the predicted probabilities for the different models provides some insight as to the doses when model form leads to the

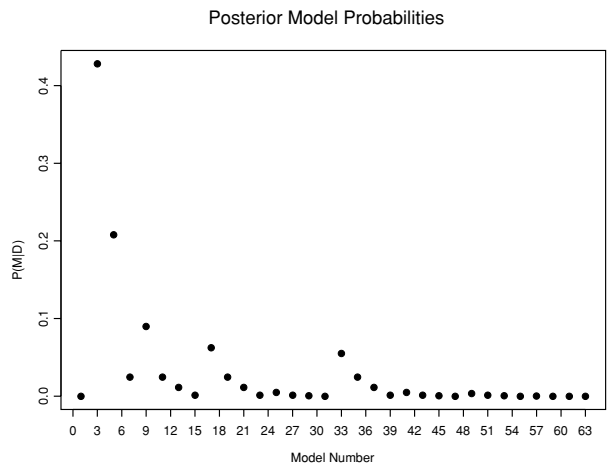


Fig. 2. Plot of posterior model probability associated with each of the 32 candidate models for multistage models fit to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene (data source: NTP (1983)).

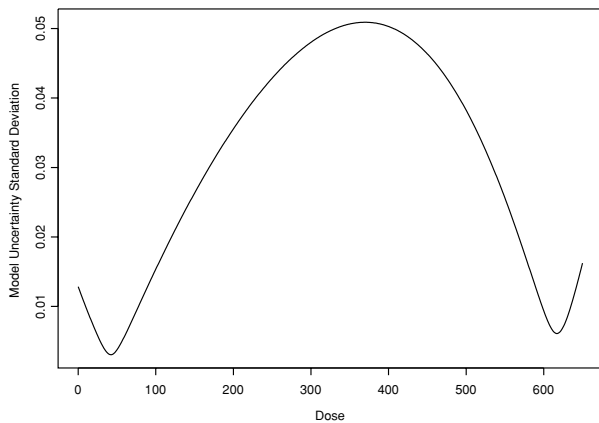


Fig. 3. Plot of standard deviation in dose-specific model predictions of tumor probabilities for different multistage models fit to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene (data source: NTP (1983)).

greatest variation in predictions. The variance component due to model uncertainty in predicted probabilities is given by $\sigma^2[\hat{Y}(d)] = \sum_{j=0}^{31} [\hat{Y}(d | M_{2j+1}) - \hat{Y}(d)]^2 \text{pr}(M_{2j+1} | D)$, where $\hat{Y}(d | M_{2j+1})$ is the predicted probability of lymphocytic malignant lymphoma for model M_{2j+1} ($j = 0, \dots, 31$) and $\hat{Y}(d)$ is the BMA-average predicted response at dose d , i.e., $\hat{Y}(d) = \sum_{j=0}^{31} \hat{Y}(d | M_{2j+1}) \text{pr}(M_{2j+1} | D)$. A plot of the model uncertainty standard deviation is plotted as a function of dose in Fig. 3. From this figure, we see that the greatest differences in model predictions are observed for a dose of approximately 350 ppm while the smallest differences in model predictions are observed for a dose near 50 or 620 ppm. The smaller model uncertainty standard deviations correspond to places where the model predictions narrow in Fig. 1.

It is interesting to compare the best model (Model ID = 3—the model with only a linear dose term) to the BMA-average predicted probabilities. This is displayed in Fig. 4. The BMA-average curve generally predicts a lower probability of lymphocytic malignant lymphoma than the best model for most doses, the BMA curve predicts a higher probability of response for the lowest doses.

Finally, we consider the BMA-average estimate of the dose associated with a specified extra risk of tumor development. Suppose that an extra risk of 1 in 1,000 is of interest, then we want to find the dose, $\Delta(d)$, that solves the equation $(\pi_i - \pi_0)/(1 - \pi_0) = 0.001$ for d . From Equation (1) (Model ID = 63), we

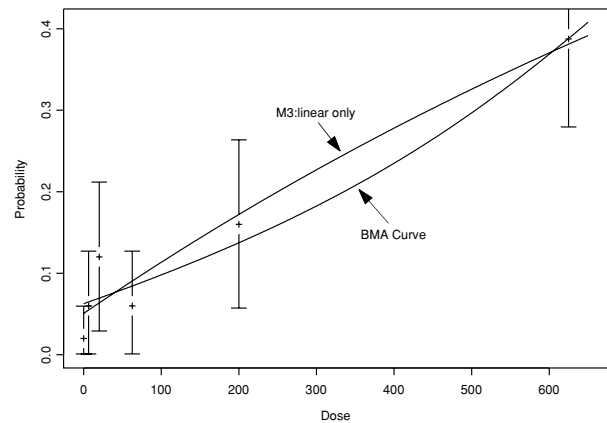


Fig. 4. Plot of estimated probability from the best fitting model (M3—only contains a linear dose term) and the BMA predicted probability based on fitting the 32 multistage models to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene (data source: NTP (1983)). The observed proportions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

have $\pi_i = 1 - \exp(-\theta_0 - \theta_1 d_i - \theta_2 d_i^2 - \theta_3 d_i^3 - \theta_4 d_i^4 - \theta_5 d_i^5)$ and $\pi_0 = 1 - \exp(-\theta_0)$. Thus, finding $\Delta(d)$ involves solving $-\theta_1 d - \theta_2 d^2 - \theta_3 d^3 - \theta_4 d^4 - \theta_5 d^5 = \ln(0.001)$ for d . Label the solution for “ d ” based on the M_{2j+1} model, $\Delta(d | M_{2j+1})$. Then, the BMA estimate of this dose is $\Delta(d) = \sum_{j=0}^{31} \Delta(d | M_{2j+1}) \times \text{pr}(M_{2j+1} | D)$. Associated with this estimate is a variance component due to model uncertainty $\sigma^2 \times [\Delta(d)] = \sum_{j=0}^{31} [\Delta(d | M_{2j+1}) - \Delta(d)]^2 \text{pr}(M_{2j+1} | D)$.

For an excess risk of 0.001, the BMA-average dose is $\Delta(d) = 36.0$ with a SD of 53.4. A plot of the estimated dose associated with an excess risk of 1 in 1,000 for each of the different models is given in Fig. 5. The radius of each plotted circle is proportional to the posterior probability of the model while the model-average estimated dose associated with the extra risk of 0.001, $\Delta(d) = 36.0$, is provided as a horizontal reference line. From this figure, we see that the estimated doses range over several orders of magnitude where the lowest doses associated with an extra risk of 0.001 are associated with models containing a linear term (Model IDs = 3, 7, 11, etc.). The posterior probability associated with selecting a model with a linear term can be determined from summing the $\text{pr}(M_n | D)$ for models containing linear terms. Here, $\text{pr}(\text{linear term in model}) = 0.4281 + 0.02451 + 0.02467 + \dots = 0.535$. Thus, the probability that a model selected from this family includes a linear term is 0.535. The largest estimated dose was associated with the most nonlinear model (Model ID = 33—containing only

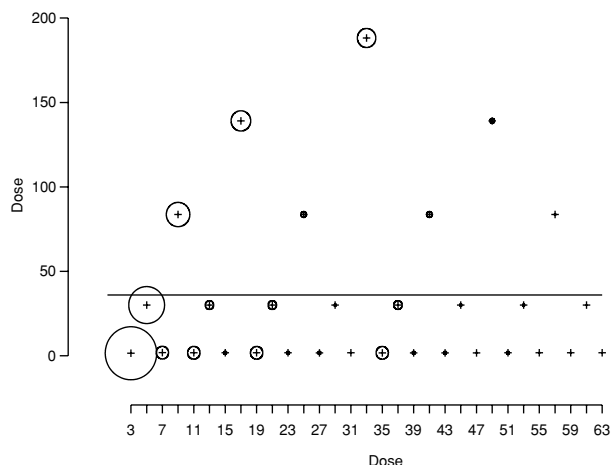


Fig. 5. Plot of the estimated dose associated with an excess risk of 1 in 1,000 for each of the different models fit to lymphocytic malignant lymphomas in female mice exposed to 1,3-butadiene (data source: NTP (1983)). The radius of each plotted circle is proportional to the posterior probability of the model. A horizontal reference line is plotted at the model-average estimated dose associated with the excess risk of 0.001.

a fifth-order term). The BMA average weights these model-specific estimates by their posterior probabilities to yield an overall model-averaged estimate. Table I also illustrates an interesting pattern where a number of models have the same posterior probability (Model IDs = 11, 19, and 35 have $\text{pr}(M_n | D) = 0.0247$) and similar estimates of the dose associated with a 0.001 extra risk (18 ppm here). These models all contain a linear term and a single higher-order dose term. These models have similar complexity in terms of numbers of parameters and provide a similar likelihood for the data.

The risk estimates are grouped in a natural way based upon model structure. The models with lowest-order polynomial term being linear yield the lowest dose estimates associated with an excess risk of 1/1,000, the models with lowest-order polynomial term being a squared dose term yield a dose estimate larger than models with the linear dose as the lowest dose term but lower than models with a cubic term as the lowest dose term and so on.

Arguably, the subset containing linear terms along with possibly higher-order terms might be of interest. If we restricted the model space to only these models, the estimated average dose associated with a 1/1,000 excess risk would be lower. This can be seen in Fig. 5. The models with linear terms are displayed as the smallest estimated doses in Fig. 5.

3.2. Example 2: BMA for BMDL of Renal Tubular Degeneration in Ethylene Glycol Exposed Rats

In the example above, we focused on a family of models that were all of the same form, the multistage quantal Model (1), but had different dose terms. This second example considers a family of binary regression models that are available in the USEPA BMD software package⁽¹⁾ for estimating benchmark doses. We considered 10 different regression models ranging from logistic to Weibull and estimated the BMD and BMDL associated with added risks of 10% or 1%. The results of this fitting exercise are given in Table II. It is worth noting that all of the models fit to these data resulted in a p -value > 0.39 for a Pearson X^2 goodness-of-fit test. The BMD10 ranged from 0.34 to 1.17% EG with BMDL10 from 0.20 to 0.70% EG, a four-fold range in BMDL. The BMD01 ranged from 0.03 to 0.67% EG with BMDL01 from 0.02 to 0.34% EG, a 17-fold range in BMDL. Thus, the estimated BMD or BMDL was more sensitive to model form for BMR = 1% as compared to BMR = 10%.

The multistage (degree = 2) Model (5) had a highest posterior probability, 0.22, and a Pearson X^2 p -value of 0.90 provided a BMDL01 of 0.03. The quantal-quadratic Model (9) had the same posterior probability, 0.22, and Pearson X^2 p -value of 0.90; however, this model provided a BMDL01 of 0.22. So, two equally attractive models in terms of goodness-of-fit or posterior model probability yield benchmark doses that differed by a factor of 7. Given no obvious preference for one model over others in this analysis, some averaging of the results would be reasonable. The BMA estimates of the BMD10 and BMDL10 were 0.83% EG and 0.45% EG, respectively, while the BMA estimates of the BMD01 and BMDL01 were 0.23% EG and 0.09% EG, respectively.

4. DISCUSSION

We have presented an illustration of incorporating model uncertainty into the risk estimation process. Note that we did not emphasize a full Bayesian analysis of the system where posterior probability distributions of the regression parameters were also of interest. Our emphasis was on the Bayesian analysis of model uncertainty to obtain a model-averaged summary. The synthesis of the results of different models has been addressed using weighting methods. For example, Buckland *et al.*⁽¹³⁾ suggested an average estimate $\hat{\theta} = \sum_k w_k \hat{\theta}_k$, where $\hat{\theta}_k$ is an estimate of the parameter of interest based on model

Table II. 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)

Model (M_m)	BMD 10%	BMDL 10%	BMD 1%	BMDL 1%	p -Value Pearson X^2	$-2\log(L)$	BIC	$\text{pr}(M_n D)$
Logistic	0.66	0.48	0.08	0.06	0.81	50.96	58.79	0.1841
Log-logistic	1.15	0.43	0.55	0.09	0.89	50.23	61.97	0.0376
Gamma	1.09	0.32	0.49	0.34	0.82	50.40	62.13	0.0345
Multistage quadratic	0.95	0.28	0.29	0.03	0.90	50.61	58.44	0.2194
Probit	0.63	0.48	0.07	0.04	0.79	51.05	58.88	0.1760
Log-probit	1.17	0.46	0.67	0.16	0.89	50.21	61.95	0.0379
Quantal-linear	0.34	0.20	0.03	0.02	0.39	53.20	61.03	0.0600
Quantal-quadratic	0.95	0.70	0.29	0.22	0.90	50.61	58.44	0.2194
Weibull	0.96	0.30	0.30	0.03	0.74	50.61	62.35	0.0311
BMA	0.83	0.45	0.23	0.09				

Note: 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 ($\text{pr}(M_n | D)$).

M_k . They suggested weights: $w_k = \exp(-I_k/2) / (\sum_{i=1}^{N_m} \exp(-I_i/2))$, where $I_i = -2\log(L) + q_i$ and N_m is the number of models considered. The adjustment term, q_i , in this information term was either $I_i = \text{AIC}$ (Aikake Information Criterion) if $q_i = 2 * (\text{no. of parameters in model } M_i)$ or $I_i = \text{BIC}$ (Bayesian Information Criterion) if $q_i = (\text{no. of parameters in model } M_i) \times \log(n)$. These authors also suggested bootstrapping with model selection as means of generating w_k . Kang *et al.*⁽¹⁴⁾ applied this with AIC-based weights to average risk or dose estimates across different microbiological dose-response models. These authors also suggested application of these ideas for the assessment of chemical risks as we conducted above. Our emphasis was on averaging results across risk estimation models for dichotomous responses where the BIC-derived weights are interpretable as posterior model probabilities.

One cost associated with this type of analysis is that now risk estimates have different values across different models and this has to be considered. This is also a benefit. 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. In this 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. Many investigators

generate risk estimates for a range of plausible models. Model averaging allows for a synthesis of risk estimates where the estimate from each model is weighted by a posterior probability of the model. While 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⁽⁷⁾ 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. (p. 784)

One strategy for using BMA in risk estimation would be to 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 this is not the case, 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.

ACKNOWLEDGMENTS

The authors thank Drs. S. Sivaganesan, J. T. Wasell, and W. K. Sieber for comments on a previous draft of this manuscript. The comments by anonymous reviewers also helped improve this presentation.

REFERENCES

1. USEPA. (2001). *Help Manual for Benchmark Dose Software, version 1.3*. EPA 600/R-00/014F. Research Triangle Park, NC: USEPA.
2. Crump, K. S., Guess, H., & Deal, K. (1977). Confidence intervals and tests of hypotheses inferred from animal carcinogenicity data. *Biometrics*, *33*, 437–451.
3. Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology*, *4*, 854–871.
4. Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A*, *158*, 419–466.
5. Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.
6. Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*, 382–417.
7. Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
8. Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
9. Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwartz criterion. *Journal of the American Statistical Association*, *90*, 928–934.
10. Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika*, *83*, 251–266.
11. National Toxicology Program. (1993). *Toxicology and Carcinogenesis Studies of 1,3-Butadiene in B6C3F1 Mice (Inhalation Studies)*. NTP TR 434. NIH Publication No. 93-3165. Research Triangle Park, NC: National Toxicology Program.
12. Robinson, M., Pond, C. L., Laurie, R. D., Bercz, J. P., Henningsen, G., & Condie, L. W. (1990). Subacute and subchronic toxicity of ethylene glycol administered in drinking water to Sprague-Dawley rats. *Drug and Chemical Toxicology*, *13*(1), 43–70.
13. Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integrated part of inference. *Biometrics*, *53*, 603–618.
14. 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.