Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/130101

Bayesian Model Averaging for Benchmark Dose Analysis in Developmental Toxicology

Eman Khorsheed[1,*] *and Mehdi Razzaghi*[2]

[1] Department of Mathematics, University of Bahrain, Sakhir, Kingdom of Bahrain
[2] Department of Mathematical and Digital Sciences, Bloomsburg University,Bloomsburg, PA, USA

Received: 2 Sep. 2018, Revised: 23 Oct. 2018, Accepted: 26 Oct. 2018
Published online: 1 Jan. 2019

Abstract: To reduce uncertainty due to model selection when a large number of potential candidate models is available, the use of Bayesian Model Averaging (BMA) has emerged as an important tool. As known, the BMA methodology is a coherent approach since we can express the desired quantities as a weighted average of model specific quantities with the weights determined based on how much the data supports each model. In toxicological studies, a wide range of statistical models have been utilized for dose-response modeling and risk assessment with no particular model receiving a universal acceptance. Here, we consider the application of BMA for benchmark dose estimation in developmental toxicity experiments. In such experiments, as in all noncancer studies, the choice of the model can play a crucial role in the final benchmark dose estimates. A Bayesian approach along with the MCMC method is used to fit each individual model used as a component in model averaging and to derive the posterior weights. A simulation study of a developmental toxicity experiment is used to illustrate the methodology.

Keywords: Model Averaging, Developmental Toxicity, Dose-Response, Benchmark Dose, Risk Assessment, MCMC

1 Introduction

The benchmark dose (BMD) approach to determine safe exposure levels in toxicological studies has become the standard and widely accepted method for risk assessment. Typically, animal bioassay experiments are conducted to assess the adverse effect of chemicals in doses that are much higher than the human exposure levels. A dose-response model is fitted to the data to establish a mathematical relationship between the response rate and the dosage level. The model is then used for extrapolation to estimate safe human exposure levels (BMDs) for a small predetermined change in response that can be considered toxicologically adverse for humans, called the benchmark risk (BMR). Technical guidelines for application of BMD approach is provided in a recent publication by the environmental protection agency [1].

A major problem in applying the BMD methodology is that many dose-response models may fit the data quite well in the experimental dose range, but when extrapolated to the low levels, the BMD estimates can often vary by an order of magnitude. To allow for more conservative considerations, often the lower confidence limit (BMDL) of the BMD is used. [2] show that ”an uncomfortably high percentage of instances can occur where the true extra risk at the BMD lower confidence limit (BMDL) under misspecified or incorrectly selected model can surpass the target BMR, exposing potential danger of traditional strategies for model selection when calculating BMDs and BMDLs”. To account for the uncertainty due to the choice of the dose-response model, model averaging has recently been introduced and utilized in a number of risk assessment problems. For example, [3] and [4] have applied the model averaging techniques for microbial risk assessment. [5] and [6] used model averaging for estimating the benchmark dose in cancer studies and found good results. In a later study, the same authors [7] used a large amount of response data to examine the performance of the model averaging procedure. Since observational data were available at the low dose region, the authors were able to estimate the benchmark dose and compare with the observed risk at low doses. They found that the model averaging technique works well and that quantitative risk estimates based on model averaging is a promising alternative to linear extrapolation based on a single model. Other authors also

* Corresponding author e-mail: ekhorsheed@uob.edu.bh

found that model averaging approach to calculating the BMDs can result in improved risk estimation [8]. In model averaging, rather than using a single mathematical relationship as dose-response model, several potential candidate models are considered and a weighted average of these models is utilized for extrapolation. An approach to model averaging that has found widespread popularity is the Bayesian Model Averaging (BMA).The advantage of BMA is that the weights are determined in such a way that they are proportional to the posterior probability that each model is correct given the observations. Therefore the weights show the extent of support in the data for each model. There was also consideration of the properties of the BMA technique in benchmark dose estimation and it was shown that the derived estimates more accurately reflect uncertainty in the understanding of the effects of exposure on the occurrence of adverse responses [9].

Here, we consider the application of BMA in quantitative risk assessment for developmental toxicity studies. In such studies, pregnant female animals are exposed to a dose of a chemical during a critical time of the gestation period. The animals are sacrificed just before term and the uterine content is examined for a variety of developmental and skeletal defects such as malformation and fetal weight. A crucial issue in modeling responses from such experiments is the consideration of the litter effect. Since it is known that responses from fetuses in the same litter behave more similarly than fetuses from different litters, incorporation of this intra-litter correlation is considered to be vital and highly important in dose-response modeling and risk assessment. Several approaches have been introduced using various techniques to analyze the data from developmental toxicity experiments. See for example [10], [11], [12], and references therein. Many of these models consider multiple outcomes. However, in the present paper in order to examine the effect of the BMA technique in developmental toxicity, we only consider a single binary outcome such as the occurrence of malformation in the fetus. A full Bayesian approach is used to fit each dose response model and the MCMC method is utilized to determine the parameter estimates for each model. In the next section, we describe our modeling approach and in section 3 our parameter estimation method is described. The application of BMA technique in developmental toxicology is discussed in section 4 and section 5 is devoted to the illustration of our methodology through simulation.

2 Model Description

Suppose that a developmental toxicity experiment consists of a control and g nonzero dose levels with $0 = d_0 < d_1 < \ldots < d_g$. Assume that m_i $(i = 0,\ldots,g)$ pregnant female dams are exposed to dose d_i according to some predetermined dose regimen. In developmental toxicity experiments, exposure generally occurs during a critical time of the gestation period. For example in experiments with mice, exposure is generally during days 5 to 12 of the gestation. Also, depending on the toxic substance, exposure could occur in a variety of formats. It could be in the diet, through gavage, by inhalation or dermal. Animals are sacrificed just before term and the fetuses are examined for developmental defects. Let n_{ij} be the number of fetuses and X_{ij} be the number of responses i.e. fetuses with a defect e.g. malformation in the j^{th} litter of the i^{th} dose level for $j = 1,\ldots,m_i$ and $i = 0,1,\ldots,g$. Then if we denote by p_{ij} the probability of response in the j^{th} litter of the i^{th} dose level, we have

$$P(X_{ij} = x_{ij}|p_{ij}) = \binom{n_{ij}}{x_{ij}} p_{ij}^{x_{ij}}(1-p_{ij})^{n_{ij}-x_{ij}} \quad x_{ij} = 0,1,\ldots,n_{ij} \tag{1}$$

Now, because of the litter effect, the probability of response p_{ij} varies from one litter to another. If we assume a beta distribution for the litter response probabilities,

$$P(p_{ij}) = B^{-1}(\alpha_i,\beta_i)p_{ij}^{\alpha_i-1}(1-p_{ij})^{\beta_i-1}, \qquad \alpha_i > 0, \beta_i > 0 \tag{2}$$

where $B(\alpha_i,\beta_i)$ is the beta function, then the marginal distribution of X_{ij} is the familiar beta-binomial model given by

$$P(x_{ij}) = \binom{n_{ij}}{x_{ij}} \frac{B(\alpha_i + x_{ij}, \beta_i + n_{ij} - x_{ij})}{B(\alpha_i,\beta_i)} \tag{3}$$

The unconditional mean and variance of X_{ij} are respectively given by

$$E(X_{ij}) = n_{ij}\mu_i \tag{4}$$

and

$$V(X_{ij}) = n_{ij}\mu_i(1-\mu_i)\{1 + \frac{\theta_i}{1+\theta_i}(n_{ij}-1)\} \tag{5}$$

where $\mu_i = \frac{\alpha_i}{\alpha_i+\beta_i}$ and $\theta_i = (\alpha_i+\beta_i)^{-1}$. Note that in this case, the intralitter correlation assumed to be the same for all litters in the same dose group is given by $\rho_i = corr(X_{ijl}, X_{ijl'}) = \frac{\theta_i}{1+\theta_i}$, where X_{ijl} denotes the pup specific response of the l^{th} pup in the j^{th} litter of the i^{th} dose group for $l = 1,\ldots,n_{ij}$, $j = 1,\ldots,m_i$, and $i = 1,\ldots,g$.

The application of beta-binomial model for developmental toxicity experiments was first suggested by Williams [13] who used it to detect a treatment effect. Expressing the model in terms of the beta distribution mean function also makes it convenient to use a link function $P(d)$ to relate the mean response to the dose through a dose-response relationship. Several dose-response models have been introduced and applied to teratological data. For example, Hoel et al.[14] used the

one-hit model. [15] considered an extension of the one-hit model that incorporated the litter effect and [16] proposed the Weibull model. Since then, there has been a variety of dose-response models that have been proposed by various authors each with some specific properties and for a review of these models, we refer to [17] and [11]. It is to be noted that the Beta-Binomial distribution is probably one of the earliest approaches for developmental toxicity modeling and once again many other approaches have been proposed. For example [18] applied the Generalized Estimating Equation (GEE) approach while [19] discussed the application of the quasi-likelihood method. Since in addition to structural malformation, in developmental toxicity studies, often other outcomes such as fetal weight, death and resorption status are also observed, several authors have developed approaches that consider multiple responses simultaneously. See for example, [20] and [21]. But the goal in this paper is not so much to compare the approaches, but rather to demonstrate the application of BMA technique in reducing uncertainty in dose-response modeling for developmental toxicity experiments. For this reason, we consider a single outcome models and we resort to the likelihood approach and apply the beta-binomial model. Now, the likelihood function is given by

$$L \propto \prod_{i=1}^{g}\prod_{j=1}^{m_i}\{\frac{\prod_{k=0}^{X_{ij}-1}(\mu_i+k\theta_i)\prod_{k=0}^{n_{ij}-X_{ij}-1}(1-\mu_i+k\theta_i)}{\prod_{k=0}^{n_{ij}-1}(1+k\theta_i)}\} \tag{6}$$

where μ_i is replaced by the probability of response $P(d_i)$ at dose d_i with $P(d)$ being a monotonic dose-response model. Once the model parameters are estimated, the process of risk assessment and determination of BMD can begin. Here, we use the additional risk, defined as the excess risk over background due to exposure i.e. $\pi(d) = P(d) - P(0)$, where $P(0)$ is the risk of an adverse effect at the background level, as the measure of increased risk. Thus if π^* denotes the BMR, a low fixed level of risk, then the BMD is the dose level corresponding to the risk $P(0) + \pi^*$. For practical purposes, the value of π^* is generally chosen to be between 0.01 or 0.1. This methodology is widely utilized to determine safe exposure levels, but unfortunately, the value of BMD can severely change depending on the choice of $P(d)$ in (6). There are several candidate models, but as pointed out in [22], "the misspecification of the risk model can adversely affect the inference on the BMD and the associated risk" This justifies the use of model averaging to reduce uncertainty in the choice of risk models which we adopt here for developmental toxicity experiments. But first, in the next section, we describe our parameter estimation method for each model.

3 Parameter Estimation

To fit each dose-response model, a Bayesian approach is adopted and the MCMC method is applied for simulation and to determine the parameter estimates. Dunson [23] points out several advantages in using a Bayesian approach to joint modeling of clustered multiple outcomes of different types. A general framework is developed using latent variables for discrete outcomes and the exact posterior distributions of parameters and latent variables are derived using MCMC methods. The approach is based on an earlier work by [24]. Bayesian approach for joint modeling of clustered outcomes has also been discussed in [25] and [26]. More recently, Bowman & George [27] developed a Bayesian methodology for joint regression modeling of discrete and continuous outcomes. Their method is assuming Gaussian latent variables for binary and ordinal outcomes and Gaussian distributions for continuous observations.

Shao & Small [28] describe a methodology for MCMC which they call a hybrid approach of Metropolis-within Gibbs-algorithm. In this approach, new samples are proposed via one parameter at a time, but with a common distribution. The advantage of this approach is that knowledge of the conditional parameter distribution is not required. New samples are kept or rejected in favor of the old one based on the ratio of prior-likelihood product of the new sample as compared to the old one. Their methodology was successfully applied in model averaging to demonstrate the reduction in the value of uncertainty in BMD estimation of carcinogenic substances when additional dose levels can be made available. Accordingly, if K candidate models are used in the model averaging procedure and if we denote the k^{th} model by $P_k(d) = F_k(\gamma_{1k} + \gamma_{2k}d^h)$ for $k = 1,\ldots,K$, then the joint posterior distribution of all model parameters given the data is proportional to the product of the likelihood and the joint prior distribution of all the parameters, that is

$$P(\gamma_1,\gamma_2|X_1,\ldots,X_g) \propto P(\gamma_1,\gamma_2) * L$$

where $\gamma_1 = (\gamma_{11},\ldots,\gamma_{1K})^T$, $\gamma_2 = (\gamma_{21},\ldots,\gamma_{2K})^T$,$X_i = (X_{i1},\ldots,X_{im_i})^T$ for $i = 1,\ldots,g$, $P(\gamma_1,\gamma_2)$ is the joint prior distribution of γ_1 and γ_2, and L is given by (6). Following [28] we also assume that the prior distributions of parameters of all the dose-response models used for model averaging i.e. $\gamma_{11},\ldots,\gamma_{1k}$ and $\gamma_{21},\ldots,\gamma_{2k}$ are independent and noninformative and that no parametric specification of the prior distribution is required. Once the parameter estimates are determined, the BMD for each model $P_j(d)$ for a given level of risk π^* can be derived from

$$BMD_k = \left[\frac{\{F_k^{-1}(\pi^*) - \gamma_{1j}\}}{\gamma_{2k}}\right]^{\frac{1}{h}} \qquad k = 1,\ldots,K \tag{7}$$

Table 1: Parameter estimates for the five dose-response models with their corresponding standard deviations in parentheses at the different BMR levels starting from 0.01.

Model	Parameter	0.01	0.025	0.05	0.075	0.1
Logistic	γ_{11}	-0.8244(0.14)	-0.8329(0.14)	-0.8163(0.14)	-0.8038(0.15)	-0.8287(0.15)
	γ_{21}	3.5041(0.75)	3.5068(0.77)	3.4523(0.81)	3.41(0.79)	3.4257(0.85)
Probit	γ_{12}	-0.1081(0.057)	-0.1444(0.09)	-0.1179(0.059)	-0.1573(0.069)	-0.1333(0.053)
	γ_{22}	0.7447(0.38)	1.0667(0.5)	0.7766(0.38)	1.1(0.63)	0.7215(0.34)
Quantal Quadratic	γ_{13}	0.1946(0.026)	0.1914(0.02)	0.204(0.026)	0.1905(0.03)	0.1962(0.027)
	γ_{23}	15.675(2.42)	17.659(3.39)	17.33(3.33)	16.1(2.34)	16.49(2.55)
Weibull ($h=5$)	γ_{14}	0.2885(0.0337)	0.2811(0.0333)	0.2763(0.0347)	0.278(0.0315)	0.2786(0.0361)
	γ_{24}	323.84(69.57)	334.05(72.71)	318.56(67.67)	337.07(62.32)	322.51(71.28)
Weibull (h unfixed)	γ_{15}	0.1679(0.025)	0.1675(0.027)	0.1619(0.029)	0.1638(0.026)	0.1735(0.028)
	γ_{25}	8.3453(3.41)	8.0121(3.74)	7.763(2.37)	7.734(2.48)	9.176(3.22)
	h	1.5291(0.21)	1.491(0.24)	1.5092(0.18)	1.4826(0.189)	1.586(0.21)

Table 2: BMD estimates for the five dose-response models.

Model	0.01	0.025	0.05	0.075	0.1
Logistic	0.0141	0.0347	0.0695	0.1051	0.1410
Probit	0.0683	0.1186	0.2899	0.4420	0.7203
Quantal Quadratic	0.0282	0.0424	0.0608	0.0776	0.0894
Weibull ($h=5$)	0.1337	0.1598	0.1858	0.1995	0.2146
Weibull (h unfixed)	0.0366	0.0615	0.0995	0.1262	0.1751

Table 3: The 5% BMDL estimates for the five dose-response models.

Model	0.01	0.025	0.05	0.075	0.1
Logistic	0.0103	0.0245	0.0486	0.0709	0.0959
Probit	0.0234	0.0370	0.0858	0.1020	0.2070
Quantal Quadratic	0.0251	0.0408	0.0530	0.0698	0.0825
Weibull ($h=5$)	0.1246	0.1493	0.1734	0.1879	0.2031
Weibull (h unfixed)	0.0139	0.0272	0.0398	0.0786	0.1363

By applying the MCMC methodology, we can generate a large sequence of parameter estimates for each model and determine the BMD. Using a measure of the center e.g. the average, we have a point estimate for BMD and using the 5^{th} percentile value, we have the corresponding BMDL.

4 Bayesian Model Averaging in Developmental Toxicology

The choice of a suitable dose-response model $P(d)$ in (6) is a critical issue and benchmark dose estimates can vary depending on the choice of the model. As pointed out in [29], it is possible to postulate different models that provide equally significant statistical fit to the data in the experimental range, but when extrapolated to low doses, give point estimates for the risk that can vary by several orders of magnitude. Using a data set from an epidemiological study, Morales et al. [30] show that in cancer studies, the risk estimate can severely vary depending on the choice of the dose-response model. For this reason, several functions have been applied in practice, but no single model has found universal acceptance for dose-response modeling in developmental toxicology. Noting this issue, Razzaghi [31] proposes the use of mixture models. The study argues that not only such models are more flexible and thus provide better fit to the data, but they also account for any non-homogeneity such as susceptibility in the population. A mixture of two logistic models was applied with some success. In this respect, therefore, the BMA methodology whereby a weighted average of several candidate models can be applied, appears to be very appealing in reducing uncertainty due to modeling. The challenge, however, to determine the weights and the attraction of the BMA technique is that the weights are determined in such a way that models which have a better fit have higher weights and conversely, models with poorer fit have a lesser contribution to the final average. For a discussion on the choice of weights and the advantages of the BMA approach, see [32].

In the BMA methodology, we begin by assuming that all the K models have a priori equal weights, that is

$$P(W_k) = \frac{1}{K} \qquad k = 1, 2, \ldots, K$$

Table 4: BIC estimates associated with the five dose-response models for varying added risk values.

Model	0.01	0.025	0.05	0.075	0.1
Logistic	1510.558	1510.278	1511.438	1513.038	1510.828
Probit	1562.318	1556.178	1561.098	1555.858	1560.244
Quantal Quadratic	1464.017	1466.998	1470.758	1466.198	1466.444
Weibull ($h=5$)	1498.038	1497.798	1499.698	1497.598	1499.082
Weibull (h unfixed)	1463.478	1463.637	1463.877	1464.357	1463.877

Table 5: Weights associated with the BIC estimates of the five dose-response models for different added risk values determined by using (10).

Model	0.01	0.025	0.05	0.075	0.1
Logistic	3.391×10^{-11}	6.278×10^{-11}	4.555×10^{-11}	1.921×10^{-11}	4.994×10^{-11}
Probit	1.953×10^{-22}	6.773×10^{-21}	7.4989×10^{-22}	9.665×10^{-21}	9.288×10^{-22}
Quantal Quadratic	0.4331	0.1571	0.0311	0.2848	0.2169
Weibull ($h=5$)	1.774×10^{-08}	3.2190×10^{-08}	1.614×10^{-08}	4.327×10^{-08}	1.775×10^{-08}
Weibull (h unfixed)	0.5669	0.8430	0.9689	0.7152	0.7831

Table 6: BMA estimates for BMD and BMDL of the five dose-response models for different added risk values obtained by using BIC weights.

BMR level	BMA for BMD	BMA for BMDL
0.01	0.03184	0.02025
0.025	0.05852	0.02934
0.05	0.09824	0.04020
0.075	0.11235	0.07608
0.1	0.15651	0.12463

Then, the final weights are determined by the posterior model probabilities, which by Bayes' theorem are given by

$$P(W_K|L) \propto K^{-1}P(L|W_K) \qquad k=1,2,\ldots,K \tag{8}$$

where $P(L|W_K)$ represents the marginal distribution of the likelihood given each model. Now, as explained in [9], the computation of the marginal distributions for calculation of the posterior model probabilities in the implementation of BMA, requires solving an integral that is difficult to calculate except for very simple cases. Indeed, in most cases, especially data related to environmental and epidemiological studies derivation of closed form solutions is not feasible and the use of the Bayesian Information Criteria (BIC) to approximate the marginal distributions has successfully been adopted. Specifically, Raftery [33] suggests the following approximation,

$$P(W_K|L) \propto exp(-\frac{1}{2}BIC(W_K)) \qquad k=1,\ldots,K$$

with

$$BIC(W_K) = -2log(\max L|W_K) + a_k log(n)$$

where a_k is the number of parameters for W_K, n is the sample size and max$L$ is the maximum of the likelihood function. Note that in developmental toxicity experiments, the sample size is the litter size. According to Wasserman [34], this approximation works well in moderate sample sizes when the covariates are independent. Thus, the weights may be determined from

$$P(W_k|L) = \frac{exp(-\frac{1}{2}BIC(W_k))}{\sum_{r=1}^{K} exp(-\frac{1}{2}BIC(W_r))} \qquad k=1,\ldots,K \tag{9}$$

This procedure has been successfully applied in several applications with dichotomous responses, see for example [35] and [22]. However, [36] suggest replacing $BIC(W_k)$ by

$$\Delta(k) = BIC(W_k) - \min_{1\le r\le K} BIC(W_r) \tag{10}$$

for calculating the weights. The advantage of this method is that the Δ values are on a continuous scale of information and are interpretable regardless of the measurement scale and whether the data are continuous, discrete or categorical. We have found that this approach is more computationally stable specially when $BIC(W_k)$ is relatively large. Using this approach on the data set utilized in [5], the same set of weights results. For more information on advantages of using the Δ values, we also refer to [37] and [38].

Table 7: DIC estimates associated with the five dose-response models for varying added risk values.

Model	0.01	0.025	0.05	0.075	0.1
Logistic	1514.71	1515.44	1514.54	1516.12	1515.08
Probit	1569.21	1564.31	1566.53	1570.00	1563.63
Quantal Quadratic	1466.39	1472.94	1475.73	1471.51	1470.13
Weibull ($h=5$)	1501.33	1501.99	1504.24	1501.05	1504.68
Weibull (h unfixed)	1467.38	1466.54	1471.30	1466.45	1465.6

Table 8: Weights associated with the DIC estimates of the five dose-response models for different added risk values determined by using (10).

Model	0.01	0.025	0.05	0.075	0.1
Logistic	1.999×10^{-11}	2.318×10^{-11}	3.639×10^{-10}	1.517×10^{-11}	1.595×10^{-11}
Probit	2.926×10^{-23}	5.649×10^{-22}	1.871×10^{-21}	3.032×10^{-23}	4.575×10^{-22}
Quantal Quadratic	0.6213	0.0391	0.0974	0.0739	0.09183
Weibull ($h=5$)	1.608×10^{-08}	1.919×10^{-08}	6.292×10^{-08}	2.853×10^{-08}	2.901×10^{-09}
Weibull (h unfixed)	0.3787	0.9609	0.9026	0.9262	0.9082

Table 9: BMA estimates for BMD and BMDL of the five dose response models for different added risk values obtained by using DIC weights.

BMR level	BMA for BMD	BMA for BMDL
0.01	0.03138	0.02086
0.025	0.06078	0.02773
0.05	0.09568	0.04108
0.075	0.12261	0.07792
0.1	0.16723	0.13136

5 Simulation

In a study of the properties of BMDL using BMA, Wheeler & Bailer [5] conclude that model averaging accounts for uncertainty in model selection and results in BMDL estimates with near nominal coverage in many situations. They suggest a list of 10 commonly-used dose-response functions all of which are in the US EPA Benchmark Dose Software [39]. [9] also use 10 functions many of which are similar to those utilized by [5]. For the purpose of the current study, we used $K=5$ dose-response models. Below is the list of these models with their respective expressions for the BMD derived from (7).

1.Logistic

$$P_1(d)=\{1+exp[-(\gamma_{11}+\gamma_{21}d)]\}^{-1}$$
$$BMD_1=\frac{1}{\gamma_{21}}log\Big(\frac{exp(-\gamma_{11})\pi^*}{1-\pi^*}\Big) \qquad (11)$$

2.Probit

$$P_2(d)=\phi(\gamma_{12}+\gamma_{22}d)$$
$$BMD_2=\frac{\phi^{-1}(\pi^*)-\gamma_{12}}{\gamma_{22}} \qquad (12)$$

3.Quantal Quadratic:

$$P_3(d)=1-exp(-(\gamma_{13}+\gamma_{23}d^2))$$
$$BMD_3=\Big[\frac{log(1-\pi^*)^{-1}-\gamma_{13}}{\gamma_{23}}\Big]^{\frac{1}{2}}$$

4.Weibull, shape parameter $h=5$:

$$P_4(d)=1-exp(-(\gamma_{14}+\gamma_{24}d^5))$$
$$BMD_4=\Big[\frac{log(1-\pi^*)^{-1}-\gamma_{14}}{\gamma_{24}}\Big]^{\frac{1}{5}} \qquad (13)$$

5.Weibull, shape parameter unknown:

$$P_5(d)=1-exp(-(\gamma_{15}+\gamma_{25}d^h))$$
$$BMD_5=\Big[\frac{log(1-\pi^*)^{-1}-\gamma_{15}}{\gamma_{25}}\Big]^{\frac{1}{h}} \qquad (14)$$

We simulate a developmental toxicity experiment which mimics a study on the effect of exposure to diethylhexyl phthaliate (DEHP) in mice. For detail of the study, we refer to [40]. We consider an experiment with dose levels 0, .044, .091, .191, and .292 g/kg of body weight. The litter sizes are generated using the empirical distribution of the number of implants from the data set.

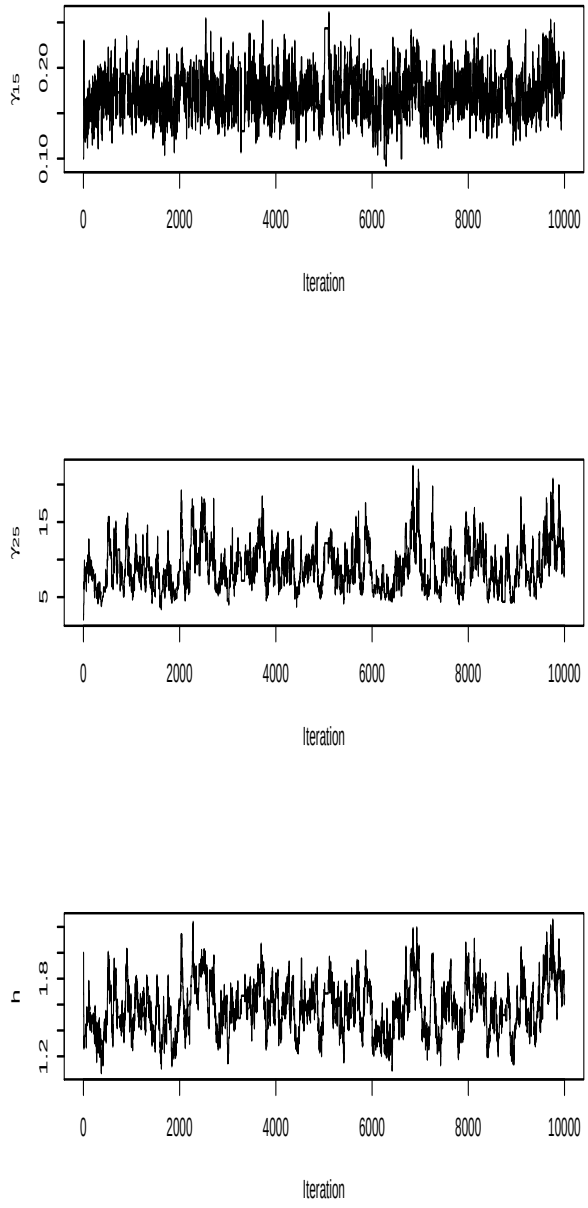

Fig. 1: Trace plots of the Weibull model parameter estimates with (h unfixed) obtained from the MCMC simulation at BMR level of 0.1. The posterior mean estimates with the corresponding standard deviations obtained after a burn-in period of 5000 iterations are: $\gamma_{15} = 0.174(0.027)$, $\gamma_{25} = 9.176(3.22)$, and $h = 1.5868(0.21)$.

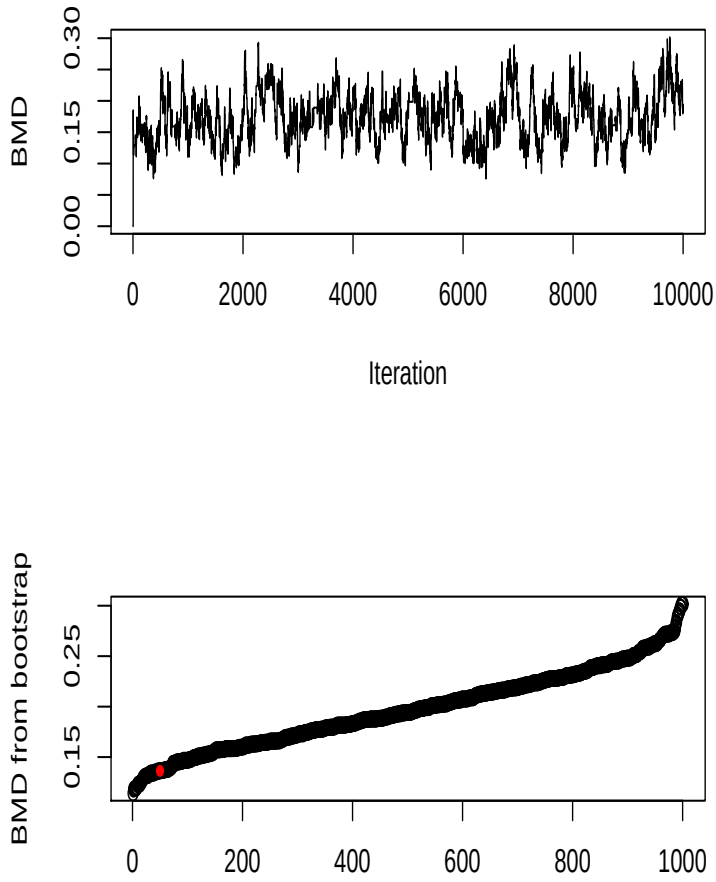

Fig. 2: MCMC trace plot of the BMD (upper panel) when the Weibull model is used with h (unfixed) and BMR level of 0.1. The BMD posterior mean is 0.1751 with standard deviation of 0.039 obtained after a burn-in period of 5000 iterations. The red dot in the 2^{nd} plot represent the 5% lower confidence bound (BMDL), which is approximately 0.1363 and is derived by sorting the last 1000 iterations of the BMD trace plot above and taking its 5th percentile.

In order to generate the number of responses per litter, we use the beta-binomial model with the probability of response being the Weibull distribution. Chen & Kodell [16] analyzed this model and used the same data set to demonstrate their methodology. Thus using the results of their parameter estimates at each dose level, we estimate the response probabilities p_{ij} using a beta distribution. The individual pup responses for the j^{th} litter of the i^{th} dose level are then determined by generating n_{ij} Bernoulli random variables with success probability of p_{ij}. For each

of the models $P_1(d),\ldots,P_5(d)$ listed above, a full Bayesian approach along with MCMC techniques based on implementation of the Metropolis-Hastings algorithm (See [41]) is used to estimate the parameters and the benchmark dose value. The BMDL is approximated via the lower 5^{th} percentile from BMD Monte Carlo chain. All simulations and computations are performed by using the statistical software package R.

6 Results

Table 1 gives the parameter estimates for the five dose-response models. Tables 2 and 3 respectively present the estimated BMDs and BMDLs at different BMR values ranging from .01 to .1. As expected, the BMDs and BMDLs increase in value as the BMR is enlarged. The BIC values and the respective model weights are displayed in Tables 4 and 5. It is clear that, as stated before, the value of BMD varies substantially based on the selected model, justifying the use of model averaging. Interestingly, the weights appear to be rather consistent and do not vary much as the risk level changes. It is encouraging to see that the Weibull model with the estimated shape parameter of approximately 1.5 and the quantal quadratic model account for 99% of the weights since the data are generated using the Weibull dose-response model of Chen & Kodell [16]. To demonstrate the MCMC results, the developed trace plots for the three parameters of our model 5, the Weibull model with unknown shape parameter, are displayed in Fig. 1. Similarly, in Fig. 2 the MCMC trace plot for the BMD resulting from model 5 at BMR level 0.1 is shown together with respective BMDLs. Finally, in order to confirm the results obtained from the BMA methodology, the alternative method of Deviance Information Criterion (DIC) for model selection is also applied to the same data. Similar to the BIC approach, the DIC approach is based on selecting a model with the lowest deviance, see [42]. Interestingly, we find that results are very similar and the model weights as well as the BMDs and BMDLs are very close to those derived from the BIC method. Tables 7-9 give the DIC values, and the BMD and BMDL values derived from applying the DIC criterion for model selection.

7 Perspective

Although risk assessment procedures and determination of safe exposure levels for noncancer endpoints have traditionally relied on the use of the No Observed Adverse Effect Level (NOAEL) and Safety Factor (SF) methods, model based approaches to determine the benchmark doses have become rather standard in recent years. Unfortunately though, estimation of BMD is generally highly model-dependent. This is the main reason why selection of the most appropriate dose-response model becomes of crucial importance. In developmental toxicology, a variety of dose response models have been utilized to derive BMD's for various toxicants with no single one being considered as the "best". In this paper, we have proposed the application of BMA methodology to reduce the uncertainty in dose-response modeling. The advantage, as demonstrated, is that models that have a low posterior probability of being correct given the data would have a lower weight. The fact that the determined weights appear to have a consistency across the varying risk levels is encouraging. In addition, the simulation results show that the procedure works well in determining the associated weight for each model. It is, however, worth mentioning that the current study considers only a single outcome. Realistically, in developmental toxicity experiments multiple outcomes are observed on each offspring. These outcomes could be a combination of discrete and continuous variables. For example, Catalano, Ryan, & Scharfstein [43] consider joint modeling of the binary outcomes fetal death and malformation while Regan & Catalano [44] discuss joint modeling of clustered binary and continuous outcomes with application in developmental toxicity studies. More recently, Najita & Catalano [45] studied the BMD determination for multiple outcomes from developmental toxicity experiments. It would be interesting to see how the BMA methodology would perform in selecting the right model or a weighted sum of models when multiple outcomes are considered.

References

[1] US EPA, Benchmark Dose Technical Guidance. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC, 2012.

[2] R. W. West, W. W. Piegorsch, E. A. Pena, W. Wu, A A. Wickens, H. Xiong, W. and Chen, The impact of model uncertainty on benchmark dose estimation, Environmetrics, Vol.23, No.8, pp.706-716 (2012).

[3] S. -H. Kang, R. L. Kodell, and J. J. Chen, Incorporating model uncertainties along with data uncertainties in microbial risk assessment, Regul Toxicol Pharmacol, Vol.32, No.1, pp.68-72 (2000).

[4] H. Moon, K. Hyun-Joo, J. J. Chen, and R. L. Kodell, Model averaging using the Kullback information criterion in estimating effective doses for microbial infection and illness, Risk Anal, Vol.25, No.5, pp.1147-1159 (2005).

[5] M. W. Wheeler, and M. J. Bailer, Properties of model-averaged BMDLs: A study of model averaging in dichotomous response risk estimation, Risk Anal, Vol. 27, No.3, pp.659-670 (2007).

[6] M. W. Wheeler, and M. J. Bailer, Bayesian monotonic semiparametric benchmark dose analysis, Risk Anal, Vol.32, No.7, pp.1207-1218 (2012).

[7] M. W. Wheeler, and M. J. Bailer, An empirical comparison of low-dose extrapolation from points of departure (PoD) compared to extrapolations based upon methods that account

for model uncertainty, Regul Toxicol Pharmacol, Vol.67, No.1, pp.75-82 (2013).
[8] W. W. Piegorsch, L. An, A. A. Wickens, R. W. West, E. A. Pena, and W. Wu, Information-theoretic model-averaged benchmark dose analysis in environmental risk assessment, Environmetrics, Vol.24, No.3, pp.143-157 (2013).
[9] M. Whitney, and L. M. Ryan, Quantifying dose-response uncertainty using Bayesian model averaging, In: Uncertainty Modeling in Dose-Response, R.M. Cooke (ed.). Wiley, N. J. Hoboken, pp.165-179 (2009).
[10] C. Faes, H. Geys, M. Aerts, and G. Molenbergs, A hierarchical modeling approach for risk assessment in developmental toxicity studies, Comput Stat Data Anal, Vol.51, No.3, pp.1848-1861 (2006).
[11] D. L. Hunt, S. N. Rai, and C. S. Li, Summary of dose-response modeling for developmental toxicity studies, Dose-Resp, Vol.6, No.4, pp.352-368 (2008).
[12] A. Kottas, and K. Fronczyk, Flexible Bayesian modelling for clustered categorical responses in developmental toxicology. In Bayesian Theory and Applications, P. Damien, P. Dellaportas, N.G. Polson and D.A. Stephens (Editors), Oxford University Press, pp.70-83 (2013).
[13] D. A. Williams, The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity, Biometrics, Vol.31, No.4, pp.949-952 (1975).
[14] D. G. Hoel, D. W. Gaylor, R. L. Kirschstein, U. Saffiotti, and M. A. Schnei- derman, Estimation of risks of irreversible, delayed Toxicity, J Toxcol Environ Health, Vol.1, No.1, pp.133-151 (1975).
[15] K. Rai, and J. Van Ryzin, A Generalized multi-hit dose-response model for low-dose extrapolation, Biometrics, Vol.37, No.2, pp.341- 352, (1981).
[16] J. J. Chen, and R. L. Kodell, Quantitative risk assessment for terotological effects, J Amer Statist Assoc, Vol.84, No.408, pp.966-971 (1989).
[17] P. L. Williams, L. M. Ryan, Dose-Response Models for Developmental Toxicology. In: Handbook of Developmental Toxicology of R.D. Hood (ed.), pp.636-666, New York: CRC Press, (1996).
[18] L. M. Ryan, Quantitative risk assessment for developmental toxicity, Biometrics, Vol.48, No.1, pp.163-174 (1992).
[19] K. -Y. Liang, and J. Hanfelt, On the use of quasi-likelihood method in tertological experiments, Biometrics, Vol.50, No.1, pp.872-880 (1994).
[20] P. J. Catalano, and L. M. Ryan, Bivariate latent variable models for clustered discrete and continuous outcomes, J Amer Statist Assoc, Vol.87, No.419, pp.651-658 (1992).
[21] H. Geys, G. Molenberghs, L. M. Ryan, Pseudolikelihood modeling of multivariate outcomes in developmental toxicology, J Amer Statist Assoc, Vol.94, No.447, pp.734-745 (1999).
[22] S. J. Simmons, C. Chen, X. Li, Y. Wang, W. W. Piegorsch, Q. Fang, B. Hu, and G. E. Dunn, Bayesian Model Averaging for benchmark dose estimation. Environ Ecol Stat, Vol.22, No.1, pp.5-16 (2015).
[23] D. B. Dunson, Bayesian latent variable models for clustered mixed outcomes, J R Stat Soc B, Vol.62, No.422, pp.355-366 (2000).
[24] J. H. Albert, S. and Chib, Bayesian analysis of binary and polychotomous response data, J Amer Statist Assoc, Vol.88, No.422, pp.669-679 (1993).
[25] J. Q. Shi, and S. Y. Lee, Bayesian sampling-based approach for factor analysis models with continuous and polytomous data, Brit J Math Stat Psycol, Vol.51, No.2, pp.233-252 (1998).
[26] D. B. Dunson, Z. Chen, and J. Harry, A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. Biometrics, Vol.59, No.3, pp.521-530 (2003).
[27] D. Bowman, and E. O. George, A Bayesian analysis of clustered discrete and continuous outcomes, J Appl Stat, Vol.45, No.3, pp.438-449 (2018).
[28] K. Shao, and M. J. Small, Potential uncertainty reduction in model-averaged benchmark dose estimation informed by an additional dose study, Risk Anal, Vol.31, No.10, pp.1561-1575 (2011).
[29] D. Krewski, D. Gaylor, and M. Szyszkowicz, A model-free approach to low-dose extrapolation, Environ Health Persp, Vol.90, pp.279-285 (1991).
[30] K. H. Morales, L. M. Ryan, T. L. Kuo, M. M. Wu, and C. J. Chen, Risk of internal cancer from arsenic in drinking water, Environ Health Persp, Vol.108, No.7, pp.655-661 (2000).
[31] M. Razzaghi, Quantitative risk assessment for developmental toxicants in non-homogeneous populations, In: Statistics for the Environment **4**: Pollution Assessment and Control, V. Barnett, A. Stein and K.F. Turkman eds, Wiley, New York, pp.209-220 (1999).
[32] K. H. Morales, J. G. Ibrahim, C. J. Chen, and L. M. Ryan, Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water, J Amer Statist Assoc, Vol.101, No.4, pp. 9-17 (2006).
[33] A. E. Raftery, Bayesian model selection in social research, Sociol Methodol, Vol.25, pp.111-163 (1995).
[34] L. Wasserman, Bayesian model selection and model averaging, J Math Psychol, Vol.44, No.1, pp.92-107 (2000).
[35] A. J. Bailer, R. B. Noble, and M. W. Wheeler, Model uncertainty and risk estimation for experimental studies of quantal responses. Risk Anal, Vol.25, No.2, pp. 291-299 (2005).
[36] K. P. Burnham, D. R. Anderson, and K. P. Huyvaert, AIC model selection and multimodel inference in behavioral ecology: some background, observations and comparisons, Behav Ecol and Sociobiol, Vol.65, No.1, pp. 23-35 (2011).
[37] D. R. Anderson, Model based inference in the life sciences: a practical information- theoretic approach, 2nd edition, Springer, New York, (2008).
[38] K. P. Burnham, and D. R. Anderson, Model selection and multimodel inference: a practical information-theoretic approach, 2nd edition, Springer, New York, (2002).
[39] US EPA, Benchmark Dose Software (BMDS). Version 2.5, http:www.epa.gov/ncea/bmds, (2014).
[40] R. W. Tyl, C. Jones-Price, M. C. Marr, and C. A. Kimmel, Teratologic Evaluation of Diethylhexyl Phthalate (CAS No. 111-81-7) in CD-1 Mice, Technical report [NCTR/NTP contract 222-80-2031, NTIS PB85105674], Springfield, VA: National Technical Information Service, (1983).
[41] W. K. Hastings, Monte Carlo sampling methods using Markov Chains and their applications, Biometrika, Vol.57, No.1, pp.97-109 (1970).
[42] T. Ando, Bayesian Model Selection and Statistical Modeling, CRC Press, (2010).
[43] P. J. Catalano, L. M. Ryan, and D. Scharfstein, Modeling fetal death and malformation in developmental toxicity studies, Risk Anal, Vol.14, No.1, pp.629-637 (1994).

[44] M. Regan, and P. J. Catalano, Likelihood models for clustered binary and continuous outcome: application to developmental toxicology, Biometrics, Vol.55, No.3, pp.760-768 (1999).
[45] J. S. Najita, and P. J. Catalano, On determining the BMD from multiple outcomes in developmental toxicity studies when one outcome is intentionally missing, Risk Anal, Vol.33, No.8, pp.1500-1509 (2013).

Eman Khorsheed is an Assistant Professor at University of Bahrain. She holds a B.Sc. in Mathematics, M.Sc. in Mathematics from the University of Bahrain, and a Ph.D. in Statistics from University of Bath, UK, which she received in 2007. She also holds a Postgraduate Certificate in Academic Practice (PCAP) from York St. John, UK, 2011. Her primary research interests include Bayesian Statistics, Markov chain Monte Carlo (MCMC) techniques, simulation, image analysis, risk analysis, Machine Learning, Data Analytics & Computational Statistics, Biostatistics, Industrial Statistics, Project Management, and Operations Research. Dr. Khorsheed has several scientific papers, reports and projects either presented at local forums, International conferences or published in scientific journals. She has served as an invited member of several International conference committees. In 2017, she was invited to join the Local Programme Committee of the International Statistical Institute (ISI) 61st World Statistics Congress (WSC) hosted by The High Commission for Planning in Marrakech. She has also chaired & organized sessions in several international conferences. In 2013, Khorsheed founded a new international journal under the umbrella of University of Bahrain: Journal of Data Analysis and Operations Research (JDAOR). Recently, she has been nominated to serve on the Editorial Board of the American Journal of Theoretical and Applied Statistics. Professional affiliations include membership with the American Statistical Association, International Statistical Institute, Tunisian Decision Aid Society, and Bahrain Society for Training & Development. Dr. Khorsheed is also a Fellow of the Higher Education Academy (HEA), UK since 2011.

Mehdi Razzaghi is Professor of Statistics at Bloomsburg University and has served in this capacity for 31 years. He holds a BS in mathematics from the University of Sussex in England, a BS in Computer Science from Bloomsburg University and a PhD in statistics from the University of London, England which he received in 1977. He held faculty positions at Marburg University in Germany, University of Kentucky, and California State University Chico, before joining Bloomsburg University faculty in 1987. Dr. Razzaghi is the recipient of the Faculty Recognition Award (2003) and the Outstanding Scholarship Award (2012) from Bloomsburg University. During the 2014-2015 academic year, he was a Fulbright fellow at the University Warsaw in Poland. In 2018, he served as a Fulbright Specialist at the Medical University of Silesia in Poland. Dr Razzaghi has collaborated extensively with the Food and Drug Administration (FDA) as a Fellow of the National Center for Toxicological Research. In his role, he assisted in the mathematical development of dose-response models and statistical risk assessment procedures in animal bioassay experiments. He further studied methods for extrapolation of the procedures for human exposure to toxic chemicals in the environment. Dr. Razzaghi has also served as a consultant with the US Environmental Protection Agency (EPA) where he was a member of the peer review panel on the effects of perchlorate environmental contamination. Further, he has consulted with the National Institutes for Health (NIH) and Geisinger Medical Center. He has been the recipient of grants from the NIH and International Life Science Institute. Professional affiliations include membership with the American Statistical Association, Society for Risk Analysis, American Mathematical Society, and International Biometric Society, among others. He is also a Fellow of the Royal Statistical Society in England.