# Web-based Bayesian Benchmark Dose (BBMD) Estimation System


**Kan Shao [1] and Andy Shapiro [2]**

[1] Assistant Professor, Department of Environmental Health, School of Public Health, Indiana University – Bloomington; [2] Independent Consultant, Durham, NC

School of Public Health Bloomington
INDIANA UNIVERSITY

Contact: Kan Shao | Email: kshao@Indiana.edu | 812-856-2725


## The Need of Bayesian BMD Estimation

The benchmark dose (BMD) methodology has mostly replaced the traditional No (or Lowest) Observed Adverse Effect Level (NOAEL/LOAEL) method in quantitative human health risk assessment. Comparing to the NOAEL/LOAEL method, the BMD method is a more quantitative dose-response modeling approach in terms of the quantitative definition of adverse effect, the derivation of the lower bound of the BMD, and etc.

There are two major BMD estimation software currently available: (1) US EPA's Benchmark Dose Software; and (2) the PROAST software published by the Netherlands' National Institute for Public Health and the Environment. Both software played an important role in facilitating and promoting the BMD methodology. The statistical methodology used in the current software is mainly based on the traditional frequentist statistics, i.e., the maximum likelihood estimation (MLE) algorithms are implemented to get point estimates for parameters of dose-response models and BMD. Although a lower bound of the BMD estimate is calculated, it is not a comprehensive description for the distribution of the BMD estimate.

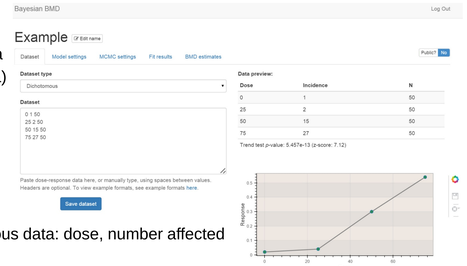
As the main stream of regulatory risk assessment moving towards a probabilistic assessment framework, a standardized tool to perform both probabilistic BMD estimation and low-dose extrapolation is of urgent need. A core idea in probabilistic dose-response analysis is that the risk of having adverse effects at a specified dose level or the dose causing a certain level of risk should be probabilistically quantified and expressed in terms of distribution. The online BMD estimation system being developed is exactly designed to achieve those goals. This system is established primarily based on Bayesian statistical analysis featuring Markov Chain Monte Carlo (MCMC) algorithms for dose-response model fitting, parameter and quantity of interest (e.g., BMD level) estimation.

## Features of the Online BBMD System

This new system includes following sections/steps: (1) data input; (2) model selection; (3) MCMC algorithms setting; (4) model fitting results; and (5) BMD estimation.

### ➢ Data Input

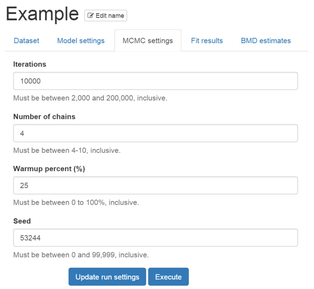
- In this step, a user selects the type of data (i.e., dichotomous data or continuous data) to be analyzed and to input the data.
- An input box is used to take data for analysis. Data can be manually input or copied and pasted from another sources.




- Three columns are required for dichotomous data: dose, number affected and number of subjects.
- Once user presses the "Save dataset" button, the dataset will be saved for future use. Additionally, a trend test will be performed to report is the data have a clear trend for dose-response modeling. Moreover, the data will be visualized (not fitted).

### ➢ Model Selection

- Seven dichotomous dose-response models currently available for selection from a pull-down menu.
- Selected models will be listed on the left and analyzed in following steps.

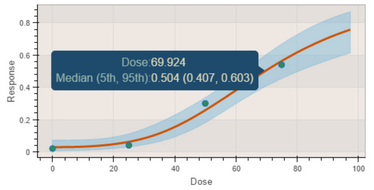




### ➢ MCMC Algorithms Setting

- The length of a MCMC chain needs to be specified; the default value is 10,000.
- The number of chains needs to be specified; the default value is 4; the maximum value is 10.
- The warmup percentage needs to be specified, the default value is 25% (i.e., the first 25% posterior sample of each chain will be discarded as burn-in).
- Random seed will be randomly generated or specified for reproduction.
- Press "Execute", then all selected model will be fit.

### ➢ Model Fitting Results

- Fitting results are available for each of the selected models. Select one particular model, the fitting results will show on the right panel, including both textual and graphical outputs.
- The first part of textual output summarizes important statistics regarding model parameters, as well as some indicators representing the quality of chain convergence.
- A parameter covariance matrix is calculated and presented.
- A distribution plot and posterior sample trace plot for each model parameter are shown.


**Posterior predictive *p*-value for model fit:** 0.516

**Model weight:** 22.6%

- Interactive DR curve is plotted, with median DR curve and 90th percentile interval are shaded.
- Moving the mouse on the curve will display the median response with 5th and 95th percentile at each dose level.
- Posterior predictive p-value (PPP) for a test statistic is calculated to indicate the adequateness of the model fitting, the p-value is calculated as follow:

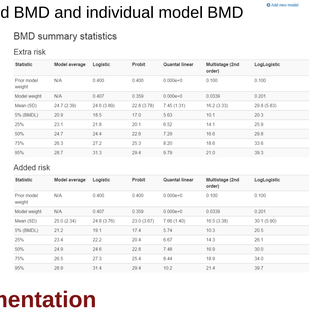
$$\Pr[T(y,\theta^t) > T(y^{rep,t},\theta^t)]$$

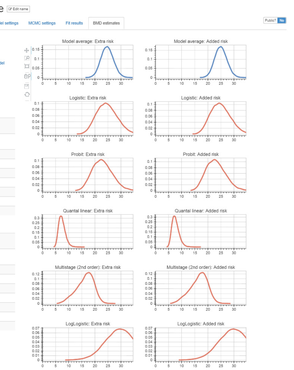
- Model posterior weight is calculated for each selected DR model for model comparison using the equation below:

$$\Pr(M_j|Data) = \frac{m_j}{\sum_{r=1}^{R} m_r} \qquad m_j = \exp(\widehat{\ell}_j - \frac{d_j}{2} logn)$$

### ➢ BMD Estimation

- A user defines a single BMR level, the BMD using both the extra- and added-risk definition are calculated
- Calculate model averaged BMD and individual model BMD
- User can specify prior model weight for selected models for model averaged BMD calculation (or remove models by specifying a prior-weight of 0)
- Textual and graphical outputs of BMD estimates are shown




### ➢ Technical Implementation

- Web-application created using the Python Django web-framework & the React.js frontend for user-interface
- Documented RESTful API can be used by other applications for batch modeling applications
- Permissions schema allows for runs to be shared publicly, should a user wish to make results public

## Improvements in the Future

### ➢ Additions in the Near Future

- Analyze continuous data for both individual and summarized data
- Allow user to specify prior distribution for parameters of dose-response models
- Calculate model averaged risk at any user-specified dose level

### ➢ Extended Implementations

- Build a prior distribution database (estimated from key studies of various chemicals) for model parameters with the potential to utilize new toxicological data