User Manual for Bayesian BMD Analysis System

Version 1.0.0.2 (2016.03.14)

This document serves as a quick guidance for using this system properly. Additionally, some statistical methods used in the system are briefly described.

In the current version, dichotomous data can be analyzed.

What's New?

- The system is now capable of analyzing continuous summary data
- The system is now capable of analyzing continuous individual data
- The system is now capable of analyzing dichotomous individual data
- The structure of the system has been improved

Contents

1. Login the System

If this is your first time use this system, you need to create an account and login to perform benchmark dose analysis by clicking "Log in" on the top right corner. Or, you can use a test-account to test the website by clicking the "Login using test-account" button.

The main benefit of creating a personal account is that your previous analyses will be saved in your account for review or update. If you login the test account, you can use this account to test all the functions but results will be deleted periodically. Your analyses results may also be modified or deleted by other users of the public account.

2. Start a New Analysis or Update an Existing Analysis

Once logged in, you reach the page with options to access previous analyses (if you have any) or to start a new analysis by pressing "New run" on the top right corner.

If you select one of your previous analyses, you can review the results in this previous analysis. You have two action options for this analysis using the pull-down menu on the upper-right corner: "update" or "delete". If you want to update an existing analysis, you basically need to use the same procedure as running a new analysis, which will be introduced in following sections.

If you select "New run", please follow the introductions below to finish a new analysis.

3. Data Input and Pre-Analysis

3.1 Selection of Data Type

In the current version, the system is capable of analyzing the following four types of dose-response data:

(1) Dichotomous summary: this type requires three values for each dose group (i.e., each input row): dose level, total number of subjects in that group, and the number of subjects affected

(2) Dichotomous individual: this type requires two values for each input row (representing each subject): dose level, and "0" or "1" indicating that the subject is non-affected or affected respectively.

(3) Continuous summary: this type requires four columns of data for each dose group: dose level, number of subjects in the dose group, mean value of the response, and standard deviation or standard error of the response.

(4) Continuous individual: this type requires two values for each input row representing each individual subject: dose level and response

The detailed differences between these data types are explained in following sections.

3.2 Data Input

(1) Once you choose "Dichotomous summary" for dichotomous summary data, three columns (from left to right) are required for input: dose level, total number of subjects and number of subjects affected. The values can be pasted or manually typed, using

spaces between values. Different dose groups should be entered in different rows. Below is an example dichotomous summary dataset:

(Dose)	(# of Subjects)	(# of Affected)
0	50	1
15.5	49	4
30	50	8
50.6	48	21

(2) If you choose "Dichotomous individual", two columns are required. The two columns from left to right are: dose and indicator (either "0" representing no effect or "1" representing with effect). Each row is used for each individual subject. An example of dichotomous individual data is shown below:

(Dose)	Indicator (0 or 1)
0	0
0	0
0	0
0	1
0	0
10	0
10	1
10	0
10	1
10	0
25	1
25	0
25	1
25	1
25	0
50	1
50	1
50	1
50	1

(3) For "Continuous summary" data type, four columns are needed to describe each dose group, which are dose, number of subjects, the mean value of response, and the standard deviation of the response (or the standard error of the response). A pull-down menu is used to specify what the last column is. The dataset below is an example of the continuous summary data:

(Dose)	# of Subjects	Mean	Std Dev
0	10	2.82	0.17
100	10	2.91	0.16
200	10	2.95	0.2
400	9	3.22	0.25

(4) The last data type is “Continuous individual” which only requires two columns (from left to right): dose level and response. The table below shows you an example of this data type.

(Dose)	(Response)
0	351.3
0	350.3
0	359.8
0	360.7
0	357.4
2.5	349.8
2.5	352.1
2.5	346.3
2.5	344.7
2.5	350.1
5	340.2
5	341.1
5	345.5
5	331.9
5	347.4
20	331.1
20	320.9
20	319.4
20	308.9
20	314.3

Once the data have been entered, press “Save dataset” to save the dataset. Please refresh your browser to make sure data set has been successfully saved.

3.3 Data Pre-Analysis

Once the dataset is saved, the “Data preview” section will display the data in a table and show the data points graphically. This plot just simply show the dose-response data but not model fitting results.

A trend test (using the Cochran-Armitage linear trend test, same as the BMDS) will be conducted once the dataset is saved. A p-value and z-score will be reported below the data table in the “Data preview” section.

4. MCMC Settings

On this page, you need to specify some settings for the MCMC algorithms.
“Iterations” is the length of MCMC chain, i.e., the number of posterior sample in each MCMC chain. Default value is 10,000. The allowable range is any integer between 2,000 and 100,000.

"Number of chains" is the number of Markov Chain to be sampled. Default value is 4. Allowable range is 4 ~ 8.

"Warmup percent (%)", the percent of sample in each Markov Chain will be discarded from the final posterior sample. Default value is 25% with a allowable range of 10% ~ 90%.

So, using the default values, the final number of posterior sample (without the warmup sample) you can get is:

$$10000 \times 4 \times (1 - 25\%) = 30000$$

"Seed" is random seed number used in the MCMC algorithms. The number is randomly generated, but you can specify the number for the purpose of reproduction.

Once these values are specified, click "Save run settings" to save the MCMC settings. Default settings are generally acceptable. However, results in the next step will provide important information that can help you judge if the MCMC settings are appropriate.

5. Model Settings

Once data input and MCMC settings are completed, go to the next tab "Model settings". In this step, you should choose the model(s) to fit the data.

5.1 Available Models and Settings

There are nine dose-response models for dichotomous data (no matter summary or individual data) and eight dose-response models for continuous data (no matter summary or individual data). Based on the data type you input and saved in the "Data input" step, a set of dose-response models for dichotomous or continuous data will show for selection.

There are nine dose-response models for dichotomous data available.

(1) Logistic model

$$f(dose) = \frac{1}{1 + e^{-a - b \times dose}}$$

Current setting for model parameter priors are:

$$a \sim Uniform(-50,50);\ b \sim Uniform(0,100)$$

(2) Probit model

$$f(dose) = \boldsymbol{\Phi}\ (a + b \times dose)$$

Current setting for model parameter priors are:

$$a \sim Uniform(-20,20);\ b \sim Uniform(0,100)$$

(3) Quantal-linear model (i.e., multistage-1st model)

$$f(dose) = a + (1-a) \times (1 - e^{-b \times dose})$$

Current setting for model parameter priors are:

$$a \sim Uniform(0,1);\ b \sim Uniform(0,100)$$

(4) Multistage-2nd model

$$f(dose) = a + (1-a) \times (1 - e^{-b \times dose - c \times dose^2})$$

Current setting for model parameter priors are:

$$a \sim Uniform(0,1);\ b \sim Uniform(0,100);\ c \sim Uniform(0,100)$$

(5) Weibull model

$$f(dose) = a + (1-a) \times (1 - e^{-c \times dose^b})$$

Current setting for model parameter priors are:

$$a \sim Uniform(0,1);\ b \sim Uniform(0,10);\ c \sim Uniform(0,100)$$

Additionally, b is limited to be larger than 1.

(6) LogLogistic model

$$f(dose) = a + \frac{1-a}{1 + e^{-c - b \times \log(dose)}}$$

Current setting for model parameter priors are:

$$a \sim Uniform(0,1);\ b \sim Uniform(0,10);\ c \sim Uniform(-50,50)$$

Additionally, b is limited to be larger than 1.

(7) LogProbit model

$$f(dose) = a + (1-a) \times \boldsymbol{\Phi}\ (c + b \times \log(dose))$$

Current setting for model parameter priors are:

$$a \sim Uniform(0,1);\ b \sim Uniform(0,10);\ \ c \sim Uniform(-50,50)$$

Additionally, b is limited to be larger than 1.

(8) Gamma model

$$f(dose) = a + (1-a) \times \boldsymbol{CumGamma}(c \times dose, b)$$

Current setting for model parameter priors are:

$$a \sim Uniform(0,1);\ b \sim Uniform(0,10);\ \ c \sim Uniform(0,100)$$

Additionally, b is limited to be larger than 1.

(9) Dichotomous Hill model

$$f(dose) = a \times g + \frac{a - a \times g}{1 + e^{-c - b \times \log(dose)}}$$

Current setting for model parameter priors are:

$$a \sim Uniform(0,1);\ g \sim Uniform(0,1);\ b \sim Uniform(0,10);\ \ c \sim Uniform(-10,10)$$

Additionally, b is limited to be larger than 1.

There are eight dose-response models for continuous data available.

(1) Linear model

$$f(dose) = a + b \times dose$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper})$$

(2) Power model

$$f(dose) = a + b \times dose^{g}$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper});\ g \sim Uniform(0,10)$$

(3) Michaelis Menten model

$$f(dose) = a + \frac{b \times dose}{c + dose}$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper});\ c \sim Uniform(0, 10)$$

(4) Hill model

$$f(dose) = a + \frac{b \times dose^{g}}{c^{g} + dose^{g}}$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper});\ c \sim Uniform(0, 10);\ g \sim Uniform(0, 10)$$

For the Linear, Power, Michaelis Menten and Hill model, the lower and upper bound of the parameter b are determined by the dose-response trend and the overall slope.
For increasing trend:

$$b_{lower} = 0$$

$$b_{upper} = \frac{Max\ Resp - Min\ Resp}{Dose_{Max\_resp} - Dose_{Min\_Resp}} \times 5$$

For decreasing trend:

$$b_{lower} = \frac{Min\ Resp - Max\ Resp}{Dose_{Min\_resp} - Dose_{Max\_Resp}} \times 5$$

$$b_{upper} = 0$$

A similar idea is applied when the input data are summary continuous data, the only difference is that the Min Resp and Max Resp will be approximated by the summary data.

(5) Exponential 2 model

$$f(dose) = a \times e^{b \times dose}$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(-20, 20)$$

(6) Exponential 3 model

$$f(dose) = a \times e^{b \times dose^g}$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(-20, 20);\ g \sim Uniform(0, 10)$$

Additionally, g is limited to be larger than 1.

(7) Exponential 4 model

$$f(dose) = a \times [c - (c-1) \times e^{b \times dose}]$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(0, 20);\ c \sim Uniform(0, 10)$$

(8) Exponential 5 model

$$f(dose) = a \times [c - (c-1) \times e^{b \times dose^g}]$$

Current setting for model parameter priors are:

$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(0, 20);\ c \sim Uniform(0, 10);\ g \sim Uniform(0, 10)$$

Additionally, g is limited to be larger than 1.

For all of the continuous models, the upper bound of the uniform distribution used for parameter a is determined in the same way:

$$a_{upper} = \max(response) \times 2$$

i.e., 2 times the largest response value in the input dataset.

Additionally, for continuous data, we need to estimate the parameter σ which is used to represent the within-dose-group variance in addition to the parameters in the dose-response model. The prior for σ is $\sigma \sim Uniform(0, 2.5)$ for all continuous models.

5.2 Add and Delete Models

Select a model from the pull-down menu then press “Create” to add the model to a list of models to be fitted, which are shown on the left panel. To add another model to the list, press “Add new model” on the left panel first, then use the same way to “Create”.

To delete or update a model, click a model name shown on the left panel, then use the “Update” or “Delete” function.

6. Execute Model Fit

When the first three tabs have been successfully completed, you can go to the fourth tab “Execute model fit”. If all of the first three steps have been saved properly, the three boxes on this page will be green. Then click “Execute” to run the program to fit the models.

After the running is finished, the fifth and sixth tab “Model fit results” and “BMD estimates” previously greyed out now is available.

7. Model Fit Results

On the “Fit results” tab, the all results obtained from the MCMC fitting process are displayed. Click one model on the left panel, then the results will be shown on the right.

7.1 Textual Output of Parameters

The first box is directly obtained from PyStan’s fit output, including some important statistics for model parameters and diagnostic indictors for the MCMC algorithms. For example, the “Rhat” can be used to judge if the MCMC chains have converged properly. If Rhat value is larger than 1.05, you may consider increase the length of MCMC chains to get better convergence. Detailed explanation on the Stan outputs can be found at: https://github.com/stan-dev/stan/releases/download/v2.9.0/stan-reference-2.9.0.pdf

7.2 Interactive DR Plot

An interactive dose-response curve is plotted below. One the plot, the median and 90th percentile interval are shown. The dose scale of the plot is from 0 to 1.5 times the maximum dose level in the input data. The dose step for the interactive DR plot is the maximum dose level on the plot divided by 1000.

7.3 Posterior Predictive P-value

A posterior predictive p-value is reported below the plot. This indicator can be used to judge if the fitting of this particular model is adequate. Practically, if the value is between 0.025 and 0.975, then the fitting is adequate. The calculation procedure is briefly described below:

(1) Use each bundle of parameters in the kept posterior sample to form a dose-response model and randomly generate case numbers, y^{rep}, at all dose levels in the original dataset
(2) Use posterior sample of model parameters to calculate a test statistic for both the original data set (d, n, y) and the replicated data set (d, n, y^{rep}). The test statistic used in this system is log-likelihood. For parameter values from the *l* th iteration, we have statistic $T(y, \theta^l)$ and $T(y^{rep}, \theta^l)$.
(3) For *l* = 1, … , L (the length of posterior sample), compare each pair of $T(y, \theta^l)$ and $T(y^{rep}, \theta^l)$, and count the number of $T(y, \theta^l) > T(y^{rep}, \theta^l)$, say *M*.
(4) The posterior predictive P-value is M/L

A detailed explanation on this procedure can be found in the Chapter of "Model checking and improvement" in *Bayesian Data Analysis* (Gelman et al).

7.4 Model Weight Calculation

A model weight is calculated and reported on this page. For each selected model, the value of $\hat{m}$ is calculated as follow:

$$\hat{m}_j = \exp\left[\hat{\boldsymbol{\ell}}_j - \frac{d_j}{2}\log(n)\right]$$

Where $\hat{\boldsymbol{\ell}}_j$ is a loglikelihood value using one set of posterior sample of parameter values, d_j is the number of parameters in the jth model, and n is the number of dose groups in the data set. So, the model weight can be calculated as (Wasserman, 2000)

$$\Pr\left(\boldsymbol{\mathcal{M}}_j \middle| Data\right) = \frac{\hat{m}_j}{\sum_{t=1}^{T} \hat{m}_t}$$

This function assumes that equal model priors for all models selected, so the weight mainly indicates how well the model fit the data. To make the weight more reliable, we used 1000 sets of randomly selected posterior sample of model parameters to calculate the model weights. And the averaged model values are reported.

7.5 Other features

Correlation matrix of the model parameters are calculated and reported.

In addition, a trace plot of the MCMC sample of the model parameters and a histogram plot with fitted density curve are also displayed.
[**Attention:** Because the trace plots are very data intense, it may cause delay and stagnation of the webpage.]

8. BMD Estimates

On this page, you can calculate the BMD estimates of your interest.

After clicking "Add new BMD", you will be able to specify some inputs for BMD calculation on the right panel.

- "BMR Name": You may make your BMD analysis profile's name identifiable, for example "BMR10% - unequal prior".
- "BMR Value": specify your BMR value, this should be a value between 0 and 1.
- "Model-weight priors": specify your model priors. By default, your selected models have an equal prior assigned automatically. If equal prior is used, then the posterior weight will be same as the model weight reported in the "Fit results" tab. The sum of the prior weights is not required to be 1, the software will calculate the correct prior model weights based on the values you input.
- Once you finish specifying all the values, click "Create" the BMD Analysis profile will be saved.
- By clicking the name of the newly created BMD Analysis profile on the left panel, the estimated results are shown on the right panel, including the distribution density plots of the model averaged BMD and single model BMD, and table with summary statistics for BMD and model weight. BMD calculated using both extra risk and added risk definitions are reported.

8.1 BMD calculation for dichotomous data

For dichotomous data, the BMD will be calculated for both the added BMR definition and the extra BMR definition. The BMDs are defined by the following equations:

Added risk:

$$f(BMD) - f(0) = BMR$$

Extra risk:

$$\frac{f(BMD) - f(0)}{1 - f(0)} = BMR$$

8.2 BMD calculation for continuous data

There are two different methods for BMD estimation for continuous data: (1) based on central tendency; (2) based on tails (i.e., the Hybrid method)

If user chooses to use central tendency to estimate BMD, then user needs to select a way to specify an adversity. There are three options:

1) Relative change:
   For this option, user need to input a value of relative change, e.g., 20%. This means that if the central tendency changes 20% from the control, it will be considered as adverse and the BMD will be calculated accordingly, using the following equation:

$$f(BMD) \pm f(0) = Relative\ Change \times f(0)$$

The plus/minus sign on the left-hand side is related to the dose-response trend, if increasing, then it is "+", otherwise it is "-".

2) Absolution change:
For this option, user need to input a value of absolute change, e.g., 3.2. This means that if the central tendency changes 3.2 from the control, it will be considered as adverse and the BMD will be calculated accordingly, using the following equation:

$$f(0) \pm Absolute\ Change = f(BMD)$$

The plus/minus sign on the left-hand side is related to the dose-response trend, if increasing, then it is "+", otherwise it is "-".

3) Cutoff:
For this option, user need to input a value of cutoff, e.g., 22.5. This means that if the central tendency is equal to the cutoff value specified, it will be considered as adverse and the BMD will be calculated accordingly, using the following equation:

$$f(BMD) = cutoff$$

The allowable range for the cutoff value will be automatically calculated based on the trend of the dose-response data and shown to user.

If user chooses to use the hybrid approach to estimate the BMD, user need to first specify an adversity value and then input a BMR value. There are two ways to specify the adversity, either specifying a cutoff value (i.e., depending on increasing or decreasing dose-response trend, above or below a value will be considered as adverse) or using a percentile value of the control (e.g., the below 1st percentile or above 99th percentile of the control distribution is considered as adverse depending on decreasing or increasing). Let's use $Q(0)$ to represent the quantile of the adversity value at control dose and use $Q(BMD)$ to represent the quantile of the adversity value at the BMD level, then the following equation should be satisfied:

For increasing trend

$$Q(0) - Q(BMD) = BMR \text{ for added risk}$$

$$\frac{Q(0)-Q(BMD)}{1-Q(0)} = BMR \text{ for extra risk}$$

For decreasing trend

$$Q(BMD) - Q(0) = BMR \text{ for added risk}$$

$$\frac{Q(BMD)-Q(0)}{Q(0)} = BMR \text{ for extra risk}$$

8.3 Posterior model weight calculation

In this step, the prior model weight specified by users will be used in the posterior model weight calculation. The function is shown below. The $\hat{m}$ for each model is calculated using the same procedure described in Section 6.4.

$$\Pr(\boldsymbol{\mathcal{M}}_j|Data) = \frac{\hat{m}_j \Pr(\boldsymbol{\mathcal{M}}_j)}{\sum_{t=1}^{T} \hat{m}_t \Pr(\boldsymbol{\mathcal{M}}_t)}$$

Based on the function above, we know that the posterior weight of a model will be 0 if the prior weight for the model is specified as 0.

8.4 Model Averaged BMD calculation

For each model, we can have posterior sample of BMD with the same length as the model parameters. Using default value, we should have:

$BMD_{1-1}, BMD_{1-2}, \dots, BMD_{1-30000}$ for model 1 with model weight W_1 (say 50%)
$BMD_{2-1}, BMD_{2-2}, \dots, BMD_{2-30000}$ for model 2 with model weight W_2 (say 30%)
$BMD_{3-1}, BMD_{3-2}, \dots, BMD_{3-30000}$ for model 3 with model weight W_3 (say 20%)

Then, the posterior sample for model averaged BMD is calculated as:

Model 1: 30000*W_1=30000*0.5, so we randomly sample 15000 values from model 1 BMD, i.e., 15000 samples from $BMD_{1-1}, BMD_{1-2}, \dots, BMD_{1-30000}$;
Model 2: 30000*W_2=30000*0.3, so we randomly sample 9000 values from model 2 BMD, i.e., 9000 samples from $BMD_{2-1}, BMD_{2-2}, \dots, BMD_{2-30000}$;
Model 3: 30000* W_3=30000*0.2, so we randomly sample 6000 values from model 3 BMD, i.e., 6000 samples from $BMD_{3-1}, BMD_{3-2}, \dots, BMD_{3-30000}$

Therefore, finally we form a new vector of model averaged BMD which also has 30,000 values, 15000 from Model 1, 9000 from Model 2, and 6000 from Model 3.

References

Wasserman, L. (2000). Bayesian model selection and model averaging. Journal of Mathematical Psychology, 44(1), 92-107.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2003). Bayesian Data Analysis. Second Edition. Boca Raton, FL: Chapman and Hall/CRC Press.