Bayesian Benchmark Dose (BBMD) Analysis System

http://benchmarkdose.com

User Manual

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. In the future the software will be exampled to also allow for continuous data analysis.

Documentation change-log

Date	Software version	Documentation changes
Dec 15, 2015	0.1.0.	First draft.

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. The main benefit of creating a personal account is that your previous analyses will be saved in your account for review or update. In addition, your runs are private (by default), but can also be made public if you wish to share.

To try the software, login using using a temporary account by clicking the "Login using test-account" button. Using the test account allows you to try the feature of the website, but results stored in the temporary account are removed daily. In addition, analyses saved in the temporary account may also be modified or deleted by other users.

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 only capable of analyzing the dichotomous data. So, choose "Dichotomous" in the "Dataset type".

3.2 Data Input

For dichotomous data, three columns (from left to right) are required: dose level, total number of subjects and number of subjects affected. The values can be pasted or manually typed, using spaces between values. Values can also be pasted directly from a Microsoft Excel spreadsheet. Different dose groups should be entered in different rows. Below is an example dichotomous dataset:

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

Once the data have been entered, press "Save dataset" to save the dataset.

Attention:

Please consecutively click "Save dataset" **twice** and then refresh this page to make sure that the dataset has been successfully saved; this bug will be fixed soon.

Once the dataset is saved, the "Data preview" section will display the data in a table and show the data points graphically.

3.3 Data Pre-Analysis

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. Model Selection and Settings

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

4.1 Available Models and Settings

There are seven 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.

4.2 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.

5. MCMC Settings

On this page, you need to specify some settings for the MCMC algorithms. "Iterations", the length of MCMC chain, i.e., the number of posterior sample in each MCMC chain. The default value is 10,000.

"Number of chains", the number of Markov Chain to be sampled. The default value is 4.

"Warmup percent (%)", the percent of sample in each Markov Chain will be discarded from the final posterior sample. The default value is 25%.

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

$$10{,}000 \times 4 \times (1 - 25\%) = 30{,}000$$

"Seed" is random seed number used in the MCMC algorithms. The default value is randomly generated, but you can specify a specific number for reproducing results.

Once these values are specified, click "Execute" to start the MCMC fitting. 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.

> **Attention:**
>
> If you selected more than five models (and used the default MCMC settings), it is likely that calculations may take a long time, and your browser may not update automatically when calculation is complete. If this occurs, please wait a few minutes, then manually refresh the page until the results are available. This. will be fixed in future versions.

6. 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.

6.1 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

6.2 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 1,000.

6.3 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).

6.4 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 1,000 sets of randomly selected posterior sample of model parameters to calculate the model weights. And the averaged model values are reported.

6.5 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:

The interactive trace plots are large and contain many values, they may render slowly and harm website performance locally on your computer; we are working to fix this issue.

7. Benchmark Dose (BMD) Calculation

On this page, you can calculate the BMD estimates of your interest. First, you need to create a profile for each of your BMD estimation analysis.

After clicking "Add new model", 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.

7.1 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 $\widehat{m}$ for each model is calculated using the same procedure described in Section 6.4.

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

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.

7.2 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:

$$BMD_{1-1}, BMD_{1-2}, \dots, BMD_{1-30000} \text{ for model 1}$$
$$BMD_{2-1}, BMD_{2-2}, \dots, BMD_{2-30000} \text{ for model 2}$$
$$\dots$$

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

$$BMD_{MA-1} = BMD_{1-1} \times w_1 + BMD_{2-1} \times w_2 + \cdots$$
$$BMD_{MA-2} = BMD_{1-2} \times w_1 + BMD_{2-2} \times w_2 + \cdots$$
$$\dots$$
$$BMD_{MA-30000} = BMD_{1-30000} \times w_1 + BMD_{2-30000} \times w_2 + \cdots$$

Therefore, we will have the same size of posterior sample for model averaged BMD. w_1, w_2, … are posterior model weight (prior model weight has been integrated) calculated from Section 7.1.

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.