参考資料 7

User Manual and Technical Guidance for The Bayesian Benchmark Dose (BBMD) Analysis System

Version 2.0 (2021.8.11)

This document primarily serves as a user manual to provide introductions for using the BBMD system properly. Technical details are also provided to help users better understand the Bayesian statistical methodology and other algorithms employed in this computational system and interpret the results.

Contents

I. OVERVIEW OF THE BBMD SYSTEM 4

A. How to Access the System 4
B. Create and Log in to an account 5
C. The User Dashboard 6
i. Access Previous Analyses 6
a. How to review a previous analysis 6
b. How to edit a previous analysis 7
c. How to share a previous analysis 8
ii. Add New Analyses 8

II. BMD ANALYSIS FOR A SINGLE DATASET 11

A. General Introduction on BMD Analysis 11
B. Data Input and Pre-analysis 12
C. MCMC Settings 15
i. How to make and change settings 16
ii. How MCMC settings may impact the results 16
D. Model settings 17
i. Dose-Response Models for Dichotomous Data 19
ii. Dose-Response Models for Continuous Data 21
iii. Dose-Response Models for Categorical Data 23
E. Model Fit Results 25
i. Parameter estimation results 26
ii. Posterior Predictive P-Value 27
iii. Posterior Model Weight 27
iv. Interactive Dose-Response Plot 28
v. Correlation Matrix 28
vi. Plots for parameter posterior sample 28
F. BMD Estimation 28
i. Dichotomous Data 28
a. How to execute a BMD Analysis 28
b. Explanation of the analysis calculation 32
ii. Continuous Data 33
a. How to execute a BMD Analysis 33
b. Explanation of the analysis calculation 37
iii. Categorical Data 38
a. How to execute a BMD Analysis 38
b. Explanation of the analysis calculation 41
iv. Model Averaged BMD Calculation 41
a. Posterior Model Weight for Model Averaged BMD Calculation 41
b. Model Averaged BMD Calculation 41
G. Probabilistic RfD Estimation 42
H. Risk/Response at Dose 44
I. Share the Analysis 46

III. BATCH PROCESSING FOR ANALYSIS 48

A. Data Input 48
B. Model Settings 49
C. MCMC Settings 50
D. BMR Settings 50

E. EXECUTE 51
F. RESULTS 51
G. SHARE/REVIEW 52

IV. BMD ANALYSIS FOR GENOMIC DATA 54

A. INTRODUCTION TO BMD ANALYSIS FOR GENOMIC DATA 54
B. DATA INPUT 55
i. How to input data into the system 55
ii. How to interpret the data input results 55
C. DATA PRE-PROCESSING 57
i. How to perform data pre-processing 57
ii. How to interpret the data pre-processing results 59
iii. Explaining the algorithms used 60
a. Fold change 60
b. P-value and adjusted P-value 61
D. DATA PREPARATION 61
E. MODEL SELECTION 63
F. BMD SETTINGS 65
i. Genomic BMD analysis steps 65
ii. Explanation of the analysis calculation 66
G. MCMC SETTINGS 66
H. BMD RESULTS 67
I. PLATFORM SELECTION 68
J. PATHWAY ANALYSIS RESULTS 69
i. How to view the pathway analysis results 69
ii. Algorithms used for pathway analysis 70
K. SHARE/REVIEW 71

V. PROBABILISTIC RFD ANALYSIS 73

A. INTRODUCTION TO THE PROBABILISTIC RFD ANALYSIS 73
B. PERFORM A PROBABILISTIC RFD ANALYSIS 73
C. INTERPRET PROBABILISTIC RFD ANALYSIS RESULTS 76

VI. BMD ANALYSIS FOR EPIDEMIOLOGICAL DATA (COMING SOON) 76

I. Overview of the BBMD System

Welcome to BBMD, the Bayesian Benchmark Dose Modeling system! BBMD is a robust tool for probabilistic dose-response assessment. Five types of dose-response analysis are available including (a) BMD analysis for single dataset, (b) batch processing for BMD analysis, (c) BMD analysis for genomic data, (d) probabilistic reference dose analysis, and (e) BMD analysis for epidemiological data (coming soon). Each type of analysis is discussed in detail from Sections 2-5. This chapter contains a description of how to assess BBMD, and an overview of the BBMD system.

A. How to Access the System

The URL of the BBMD online system is Bayesian BMD (benchmarkdose.com). Chrome and Firefox are the two recommended web-browsers for using the system. The front page of the BBMD system is displayed in Figure 1 below.

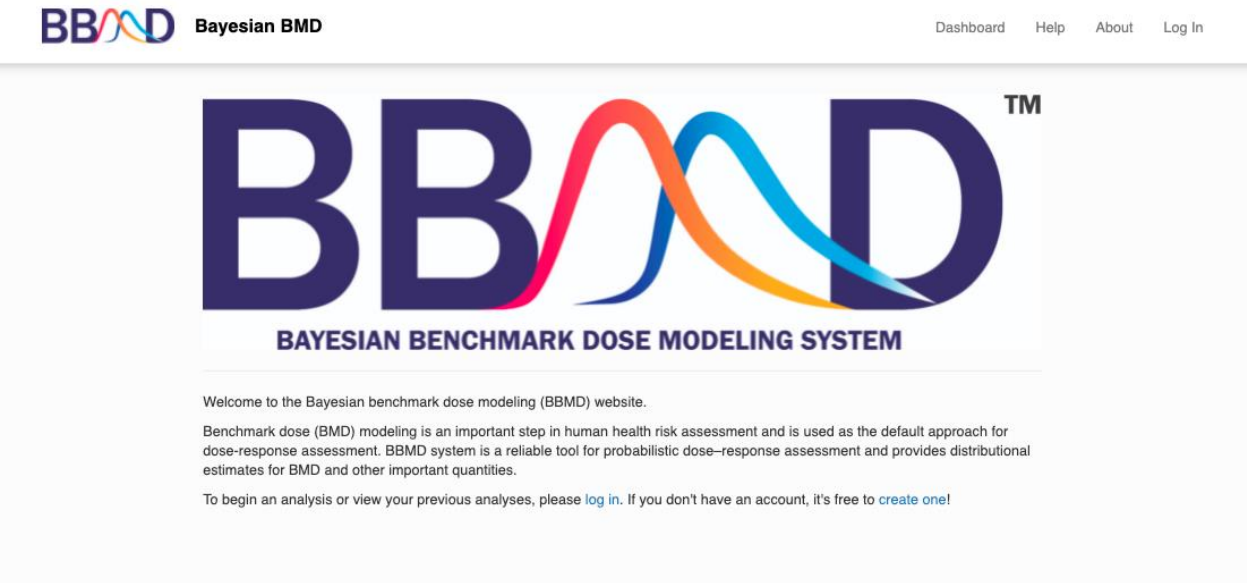


Figure 1. Front page of the BBMD system

In the upper right-hand corner of the webpage (Figure 1), there are four different options. These four options are available on every page of the system. The first option, “Dashboard”, which is where you can access your analyses. This includes analyses you have created and analyses that have been shared with you, which are stored in their respective tabs. The next option in the right corner is “Help”. This link takes you to the download of this user manual, something you must have already figured out. The third option, “About”, is a brief overview on the BBMD system, which gives you 1) a summary of this system, 2) references on the methodology of BBMD, 3) preferred citation of this website, 4) contact info of the DREAM Tech LLC development team for any questions or suggestions, and finally 5) the funding agency. The final option in the upper right-hand corner is “Log in”. To use the system for BMD analysis, users first need to either log in to the system.

B. Create and Log in to an account

By clicking the "Log-in" option, the "Log-in" page is shown in Figure 2 below. An email address (i.e., the username) and your password are required when logging in. If this is your first time using this system, you will need to create an account by clicking the "Create an account" link at the bottom of the login page shown in Figure 2.

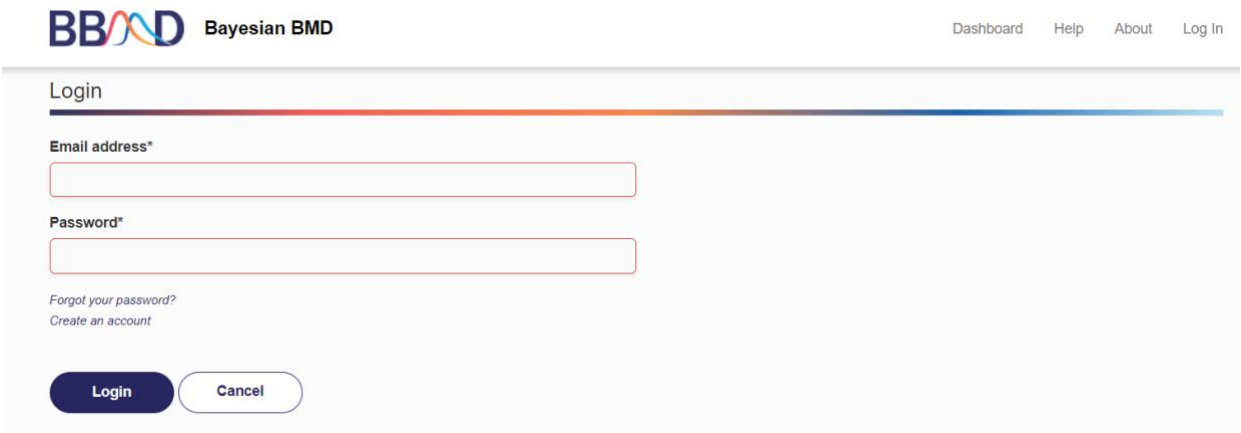


Figure 2. Login page of the BBMD system

The "Create an account" page is shown in the Figure 3. Just like most online systems, an email address (used as an account name) and password, which is entered twice, are needed to create an account. By creating a personal account, your previous analyses will be saved in your account for future review or update. A confirmation email will be sent to your inbox for verification after creating an account. Click on the link in the email to begin using the BBMD software.

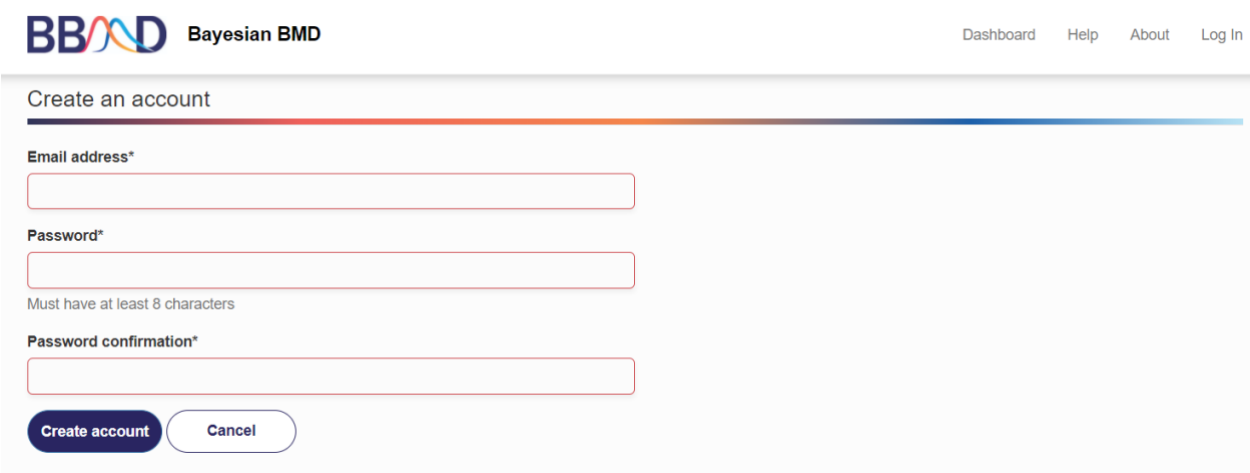


Figure 3. "Create an account" Page of the BBMD System

C. The User Dashboard

i. Access Previous Analyses

a. How to review a previous analysis

Once logged in, you will reach the page with options to access your previous analyses (if any are completed), or analyses shared with you (if you have any) or analyses you shared with the experts from Dream Tech LLC (if you have any) in the respective tabs. If you select one of your previous analyses (e.g., "EXAMPLE" in Figure 4), the previous results will be shown as the one is shown in Figure 5, you can review the previous analysis. For each type of analysis, there are unique tabs available. By clicking each of the tabs, you can access the data or settings stored in this analysis. One or more tabs may be empty, which indicates that the analysis was previously ended before the corresponding section had been finished. The detailed information regarding the contents in each of the tabs will be introduced in the following sections.

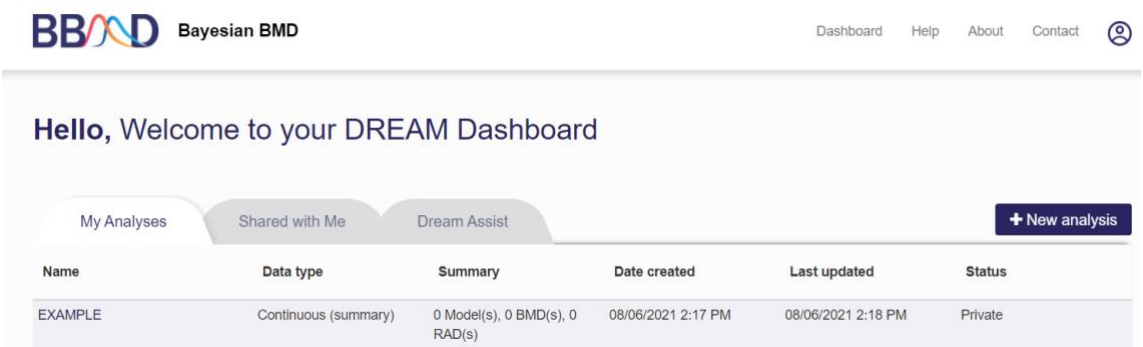


Figure 4. The Summary Page of Existing Runs

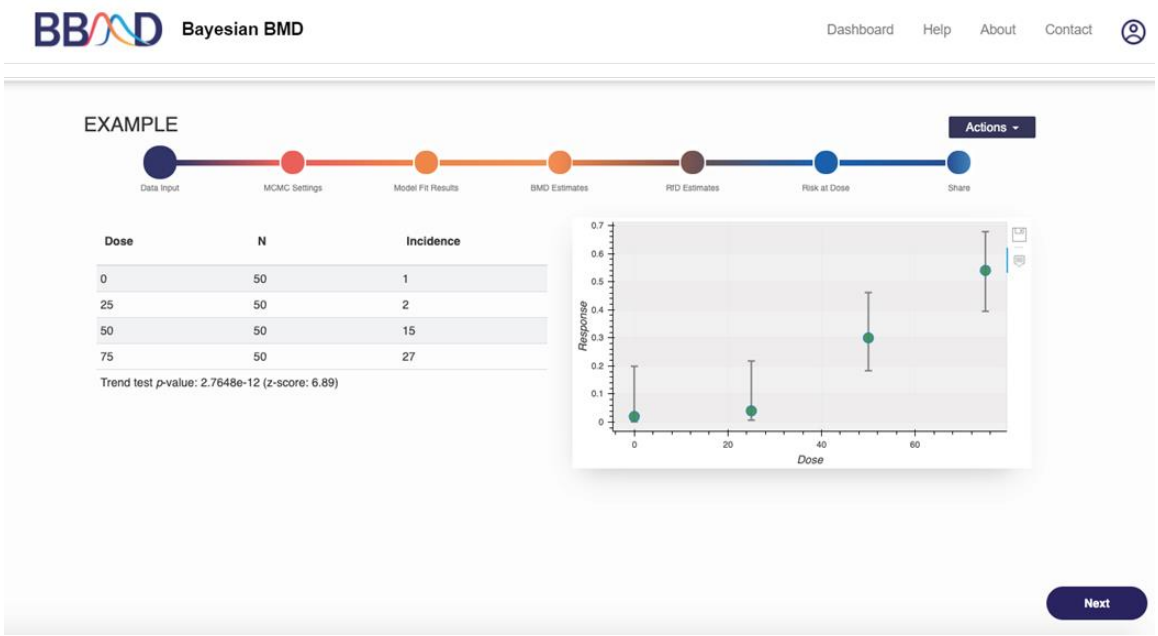


Figure 5. Reviewing an Existing Analysis

b. How to edit a previous analysis

You can also edit previous analysis. The pull-down menu of 'Action' in Figure 6 has the 'Update' and 'Delete' options. 'Update' takes you to the editing phase of an analysis session where you can specify analysis settings and execute analysis estimations. Selecting 'Delete' will prompt you to confirm you want to delete this analysis. If you choose to delete an analysis, the analysis and all results will be removed from your profile.

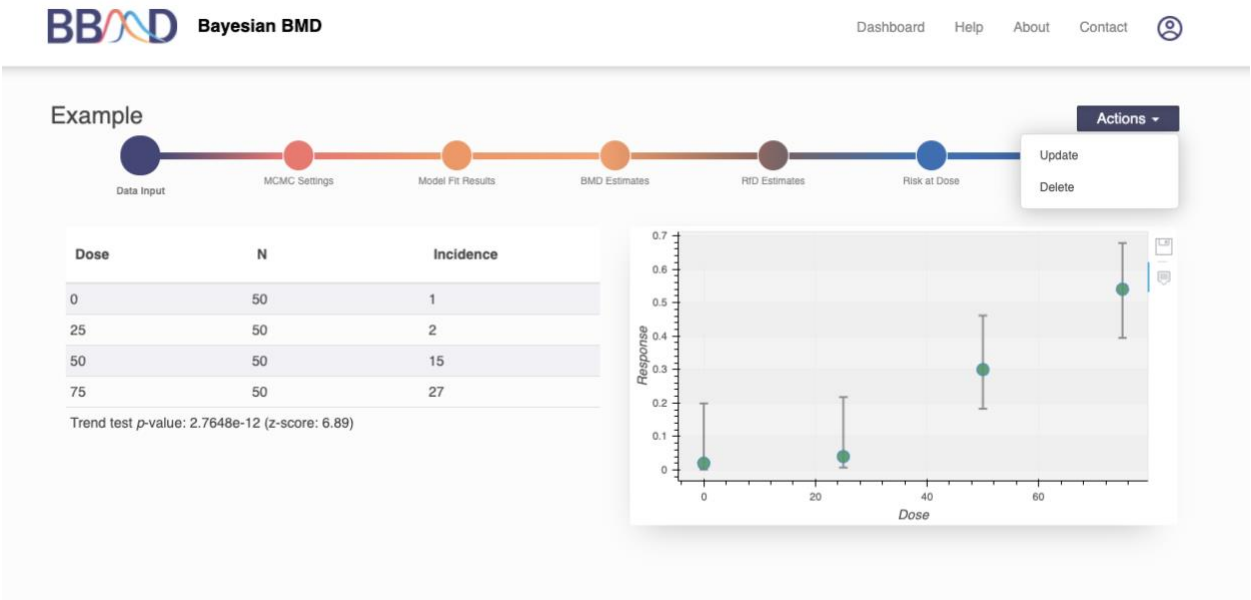


Figure 6. The Actions Available to Users for an Existing Analysis

c. <u>How to share a previous analysis</u>

The final tab in every type of analysis is the "Share" tab, shown in Figure 7. From this page you can share your analysis with the public, allow other people to edit the analysis, and even send your analysis to the DREAM Tech team who can assist with the analysis. If wish to share your analysis with the public, you will be given a share link which can be sent to different users. You can also invite users directly using their account email. The other two types of sharing, 'Organizational' and 'DREAM Assist', are coming soon and are only accessible for premium subscription accounts.

Under the "share" tab, you may also be able to export the results of the analysis into Word or Excel formats, depending on the type of analysis. Exporting the results will send a link to your account's email where you can download the reports.

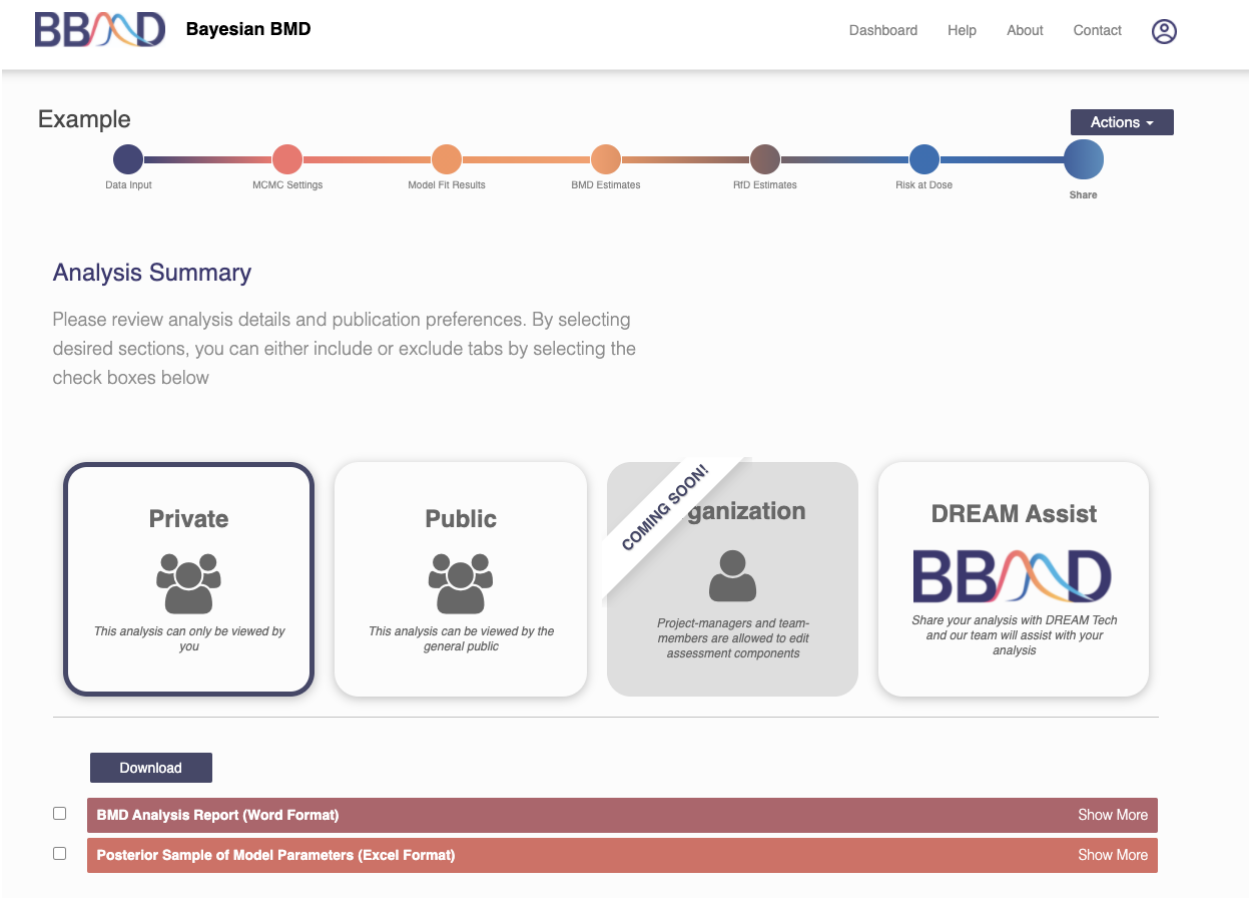


Figure 7. Sharing an analysis

ii. Add New Analyses

To start a new analysis from scratch, click “New analysis” in the top right corner in Figure 4. When starting a new analysis, you will be asked to select which type of analysis you would like to perform (Figure 8). The current analysis types are:

a) BMD Analysis for Single Dataset,
b) Batch Processing for BMD Analysis,
c) BMD Analysis for Genomic Data,
d) Probabilistic Reference Dose (RfD) Analysis,
e) BMD Analysis for Epidemiological Data (coming soon).

After specifying the data type on the analysis selection page, you will directed to the corresponding module. If the “BMD Analysis for Single Dataset” module or “Batch Processing for BMD Analysis” module is chosen, the data type either continuous or dichotomous or categorical also needs to be specified. These data types are described below. For each type of analysis, a detailed explanation of the required inputs, modeling settings, and model outputs are in sections 2-5.

- Continuous – A continuous response is reported as a measurement of the effect, such as body or tissue weights, in control and exposure groups. The response may be reported in either absolute or relative change from control. When individual data are available, the dose and response data input to BBMD for BMD inference. Instead, when individual data are not available, the summary data are needed including dose level, number of subjects in each dose group, mean value of the response, and the standard deviation or standard error of the response.
- Dichotomous - A dichotomous response is reported as either the presence or absence of an effect. Dichotomous data are reported either as the summary data with the number of animals showing the effect at each individual level or as the individual data with “0” or “1” indicating that the subject is non-affected or affected respectively at each dose level. In BBMD, dichotomous summary data require 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. Dichotomous individual data require 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.
- Categorical – A categorical response is classified as the one or more defined category (e.g., mild, moderate, or severe change) in addition to the no-effect category. This type requires three values for each input row representing each individual subject: dose, severity level, and response.

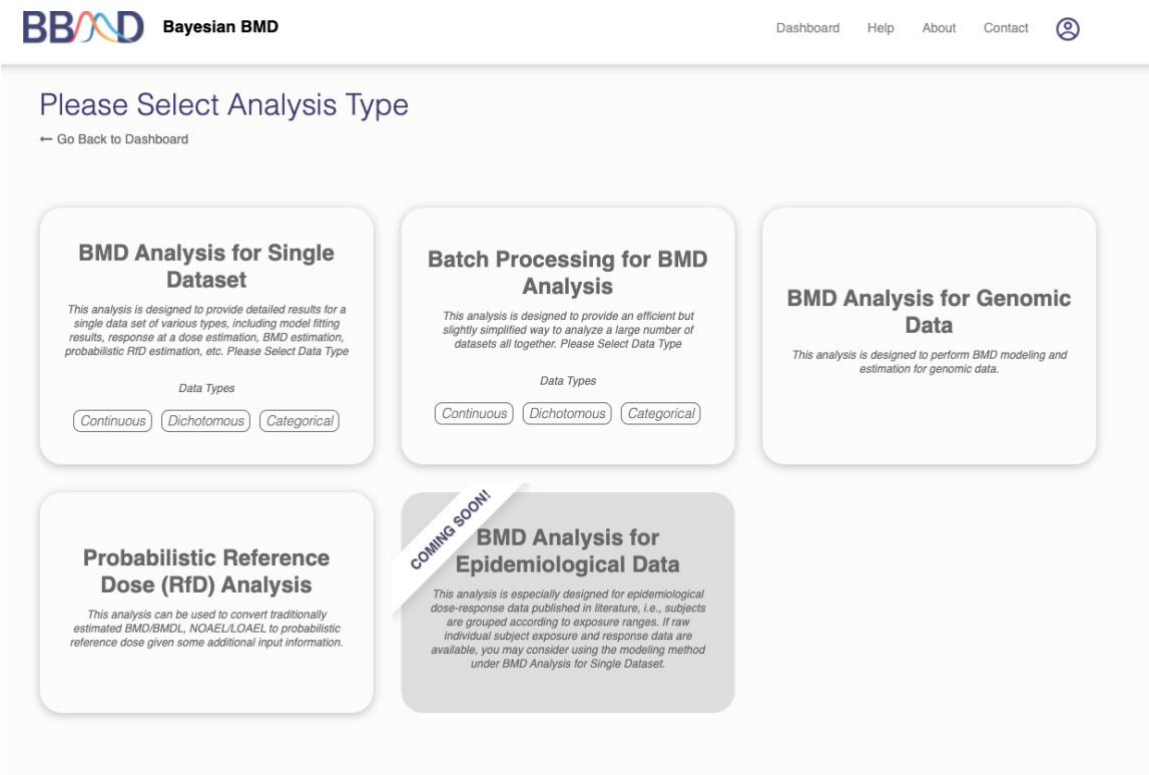


Figure 8. Analysis type selection screen when beginning a new analysis

II. BMD Analysis for a Single Dataset

A. General Introduction on BMD Analysis

After you selected 'BMD Analysis for Single Dataset' and specified the data type in Figure 8, the web page will change to the one as shown in Figure 9. An automatically generated name, "New run, *Month Day Year, HH:MM PM*", is assigned to the newly started analysis. You can click the pencil button next to the analysis name, as shown in the Figure 9, to make the name more identifiable. Without inputting any data, you will not be able to continue to advance through the tabs. The first step in an analysis is to input dose-response data. For continuous and dichotomous data, whether the dataset is 'summary' or 'individual' needs to be specified. After each step is finished, the 'Next' button and the tab in the progress bar will light up.

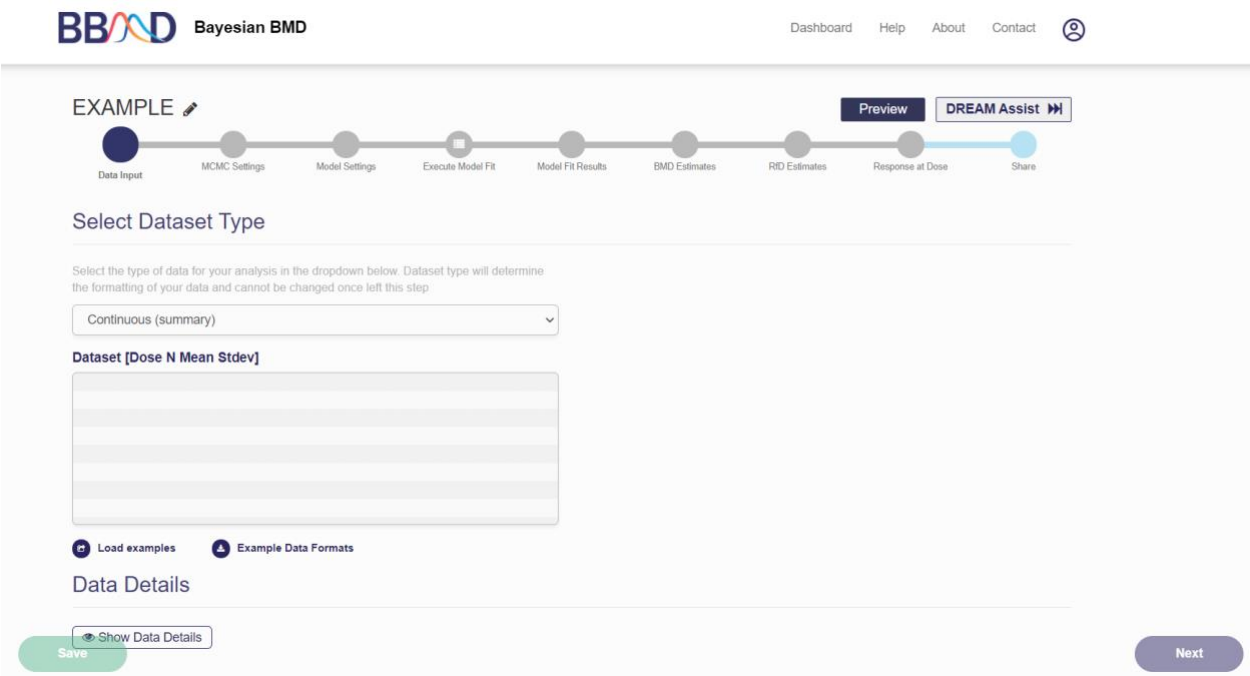


Figure 9. The Start Page of a New Analysis

As a note, if you would like to preview the results or view the results from the perspective of someone the analysis was shared with, click the blue button labeled "Preview" in the upper right corner (Figure 9). This will take you out of the updating mode. To resume updating the analysis, click the blue button labeled "Actions" in the upper right corner, and then in the drop-down list that appears, click "Update" (Figure 10).

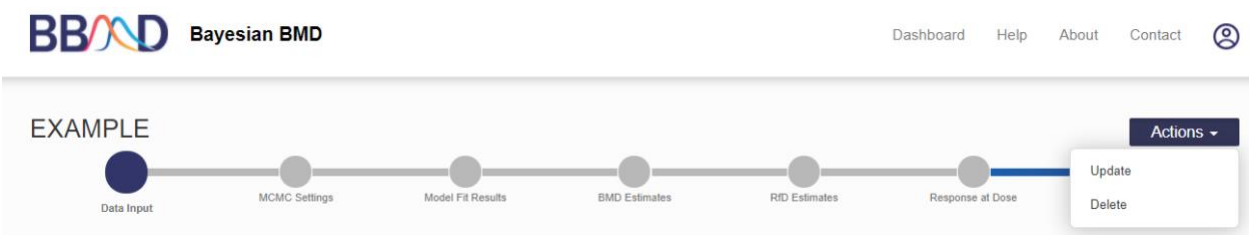


Figure 10. Preview mode

B. Data Input and Pre-analysis

The first tab, which is presented after confirming the selected data type, is the dataset input tab. Select the specific type of data from the dropdown menu, and then put the data into the text box. For continuous and dichotomous datatypes there will be two options in the dropdown menu: Either summary or individual data. If are doing an analysis with categorical data, there is only one option. Each type of data requires different columns. The following section will explain the types of columns needed for each data type.

Dichotomous summary

If you choose "Dichotomous summary" for dichotomous data, three columns are required for input (from left to right): 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)	(Number of Subjects)	(Number of Affected)
0	50	1
15.5	49	4
30	50	8
50.6	48	21

Dichotomous individual

If you choose "Dichotomous individual", two columns are required. The two columns are (from left to right): 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

Continuous summary

For "Continuous summary" data type, four columns are needed to describe each dose group: dose, number of subjects, the mean value of response, and the standard deviation of the response 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

Continuous individual

The "Continuous individual" data type 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

Categorical

The "Categorical" data type requires three columns (from left to right): dose, severity level response. The table below shows you an example of this data type

Dose	Severity	Response
0.00005	0	22
0.00005	1	0
0.00005	2	0
0.00005	3	0
0.01	0	8
0.01	1	0
0.01	2	0
0.01	3	0
0.025	0	0
0.025	1	8
0.025	2	8
0.025	3	0
0.05	0	0
0.05	1	1
0.05	2	15
0.05	3	0
0.075	0	0
0.075	1	0
0.075	2	4
0.075	3	0
0.1	0	0
0.1	1	0
0.1	2	2
0.1	3	2

Once the data have been entered, press "Save" to save the dataset. Please refresh your browser to make sure the data set has been successfully saved. Once the data set is successfully saved, the data will be visually displayed and summarized in a table as shown in the Figure 11. If you hover your mouse over marks on the dose-response plot, you can see more detailed information regarding that point on the plot.

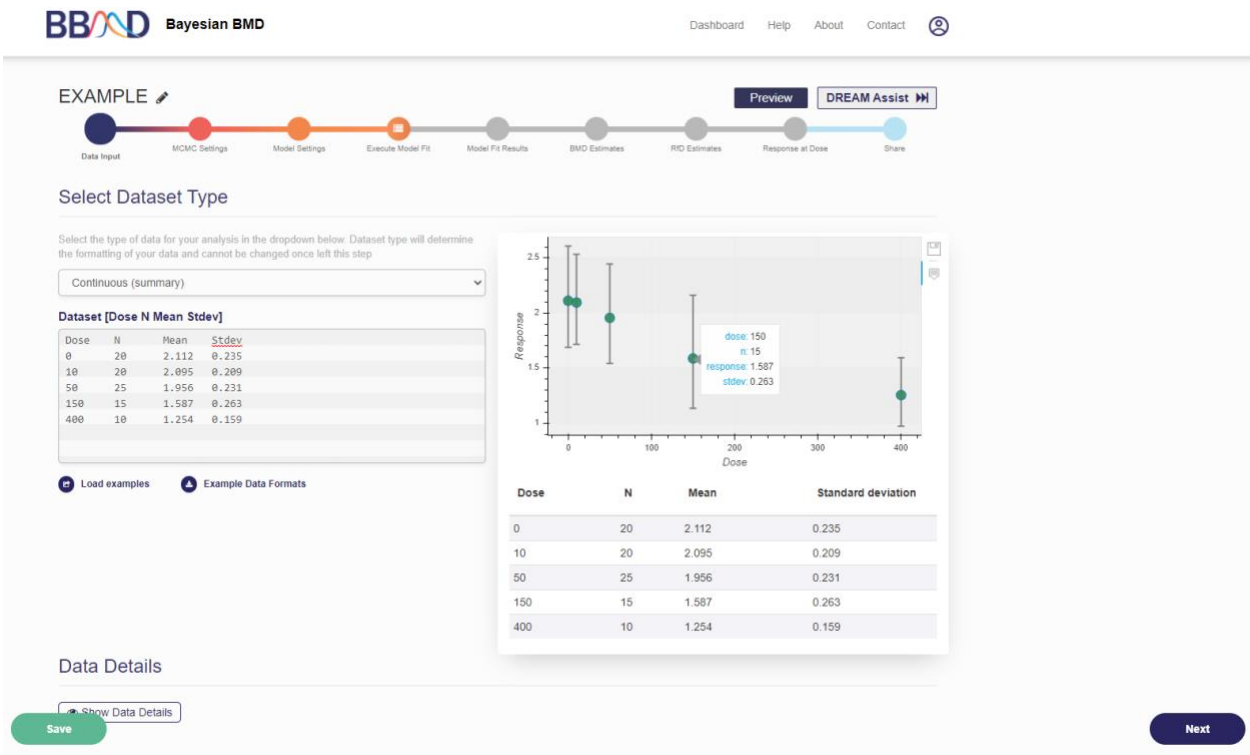


Figure 11. Data input page after inputting data

In addition to the table and the dose-response plot, if you uploaded dichotomous data, a trend test (using the Cochran-Armitage linear trend test, same as the BMDS) is performed. A p-value and z-score will be reported below the data table. The trend test results can be used to judge if the dichotomous dataset is appropriate for BMD modeling.

To continue to the next tab, press “Next” in the lower right corner.

C. MCMC Settings

On this tab (shown in Figure 12), you can specify some settings for the MCMC algorithms.

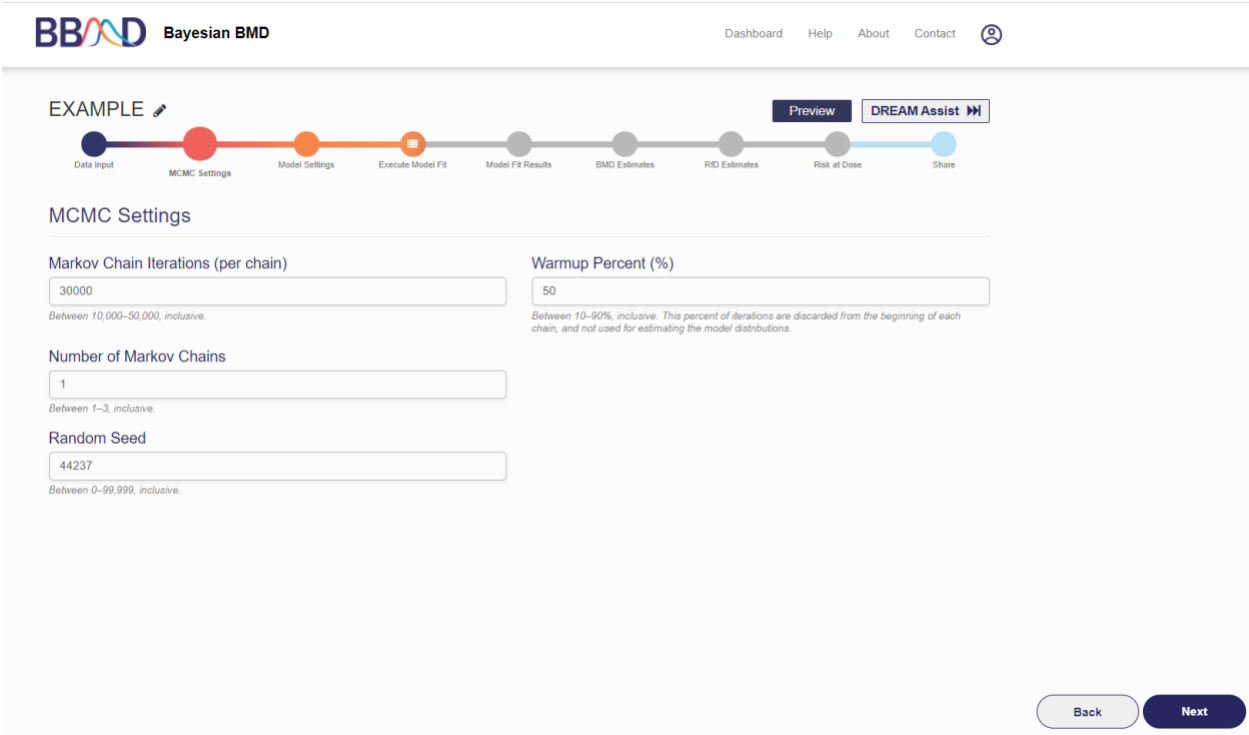


Figure 12. MCMC settings

i. How to make and change settings

There are four different values that need to be specified in this tab. First, specify the number of Markov chain iterations, between 10,000 and 50,000 (inclusive) iterations per chain. Enter your value into the "Markov Chain Iterations" text box. Next, you need to specify the warmup percentage for each Markov Chain. This is the percentage of iterations discarded from the beginning of each chain; Therefore, those iterations will not be used for estimating model distributions. Put this percentage in the "Warmup Percent (%)" text box. Third, specify the number of Markov chains used in the analysis. Enter a number 1 to 3 (inclusive) into the "Number of Markov Chains" text box. Each chain will use the number of iterations previously specified. The final value is the random seed which is used for reproducing analysis results. The random seed can be 0 to 99,999 (inclusive). Enter this value in the "Random Seed text box".

Once these values are specified, click "Next" to save the MCMC settings and move to 'Model 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. Based on our testing, **the default settings are adequate for most of the commonly seen dose-response shapes, so we suggest you use the default settings for your initial run**.

ii. How MCMC settings may impact the results

"**Iterations**" is the length of MCMC chain, i.e., the number of posterior samples in each MCMC chain. Default value is 30,000. The allowable range is any integer between 10,000 and 50,000.

"**Number of chains**" is the number of Markov Chains to be sampled. Default value is 1. Allowable range is 1 - 3.

"**Warmup percent (%)**", the percent of sample in each Markov Chain will be discarded from the final posterior sample. Default value is 50% with an allowable range of 10% - 90%.

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

$$30000 \times 1 \times (1 - 50\%) = 15000$$

"**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.

D. Model settings

Once data input and MCMC settings are completed, click "Next" to go to the next tab, "Model settings". In this step, you should choose the model(s) to fit the data. To add a dose-response model, click the plus icon and then click "Create Model" or 'Standard Model' on the left panel, shown in Figure 13. 'Standard Model' includes all default models for each type of data which are described in section i-iii.

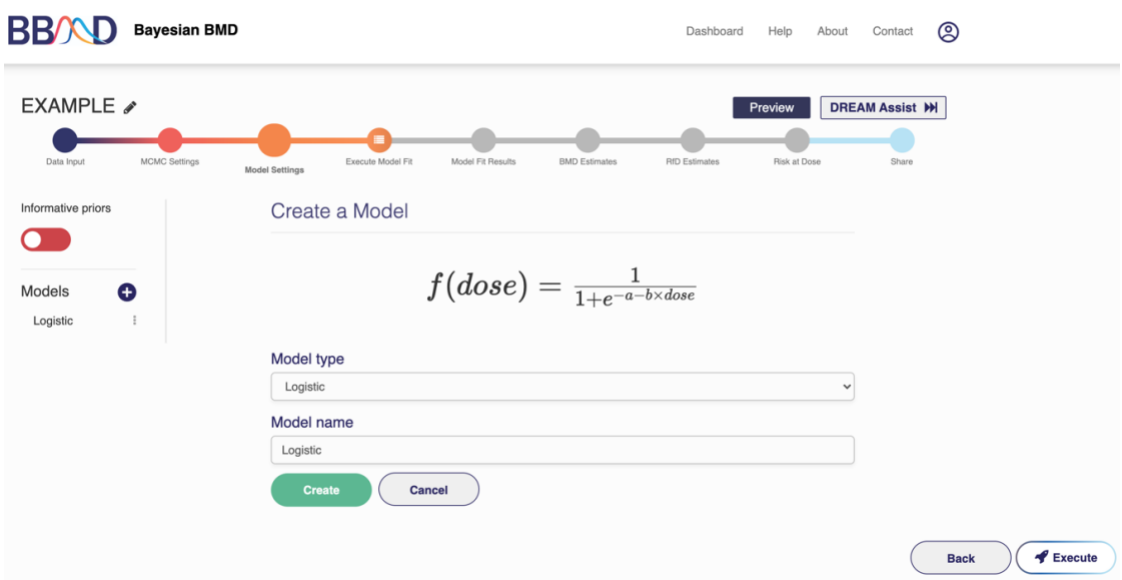


Figure 13. "Model Settings" tab

Follow these four steps to add a dose-response model to the list of models for the analysis:

1) Click the plus button in the left panel
2) Click "Create Model" in the box that appears
3) Select one model from the pull-down menu, then give an identifiable name to the model. For models with a power parameter, you need to choose a restriction value for

the power parameter. There are five options available in the current system: 0, 0.25, 0.5, 0.75, and 1. The default value is 1.

4) Press the "Create" button to add the model to the list of models on the left panel.

To add another model, repeat the steps 1) to 4). The same model with different settings (e.g., the restriction value put on the power parameter) can be added again as a separate model.

To update a model from the model list, follow these 3 steps:

1) Click the three dots next name of the model you want to update or delete in the list of models on the left panel.
2) If you want to change the current settings of the model, click "Edit Model", then you can modify the model settings like model type, model name, and the restriction value
3) Click "Update" below the settings to update the setting

You can delete a model which is already in the list of models by clicking the three dots next to the model's name, and then clicking "Delete Model" and "Delete Model" in the window that pops up (Figure 14).

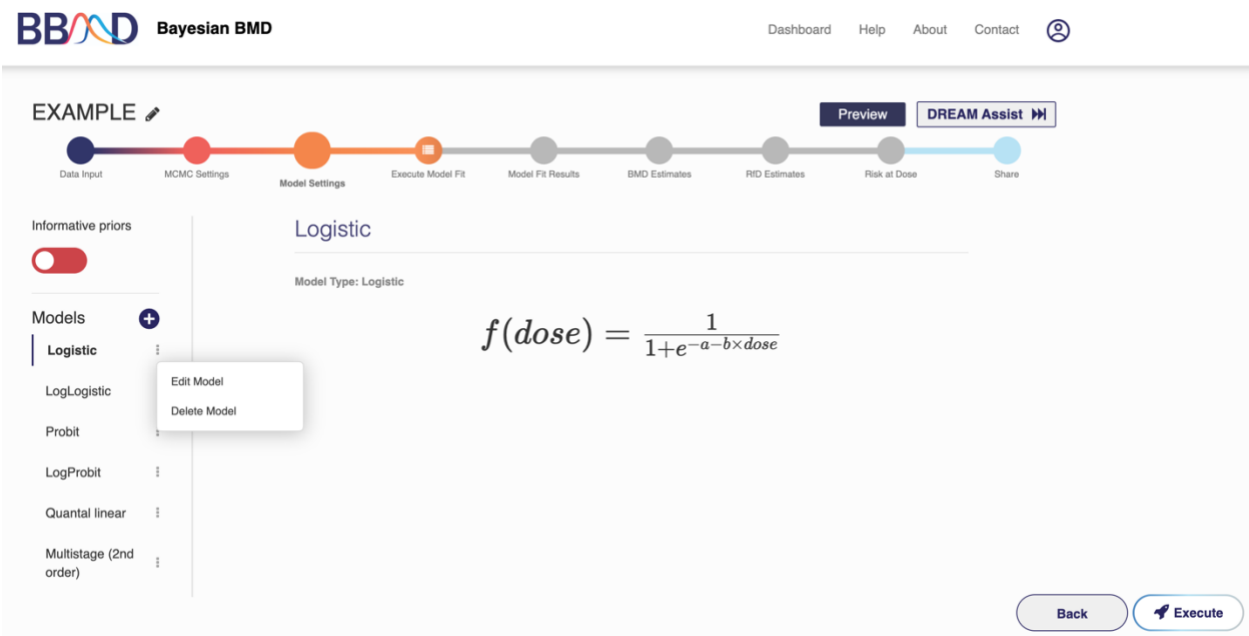


Figure 14. Update or Delete a Model in the List

Additionally, for dichotomous data, you can choose to use informative priors for the model parameters. To turn informative priors on, change the slider in the left column below "Informative Priors" to green. Then, when adding a new dose-response model, each parameter in the model complies a specified prior distribution, and you can specify the parameters of the prior distribution such as mu and sigma for a normal distribution, or beta and alpha for the beta, and gamma distributions. Once the distributions are set, distribution curves for the parameters will also be displayed with the dose-response model formula, as shown in Figure 15.

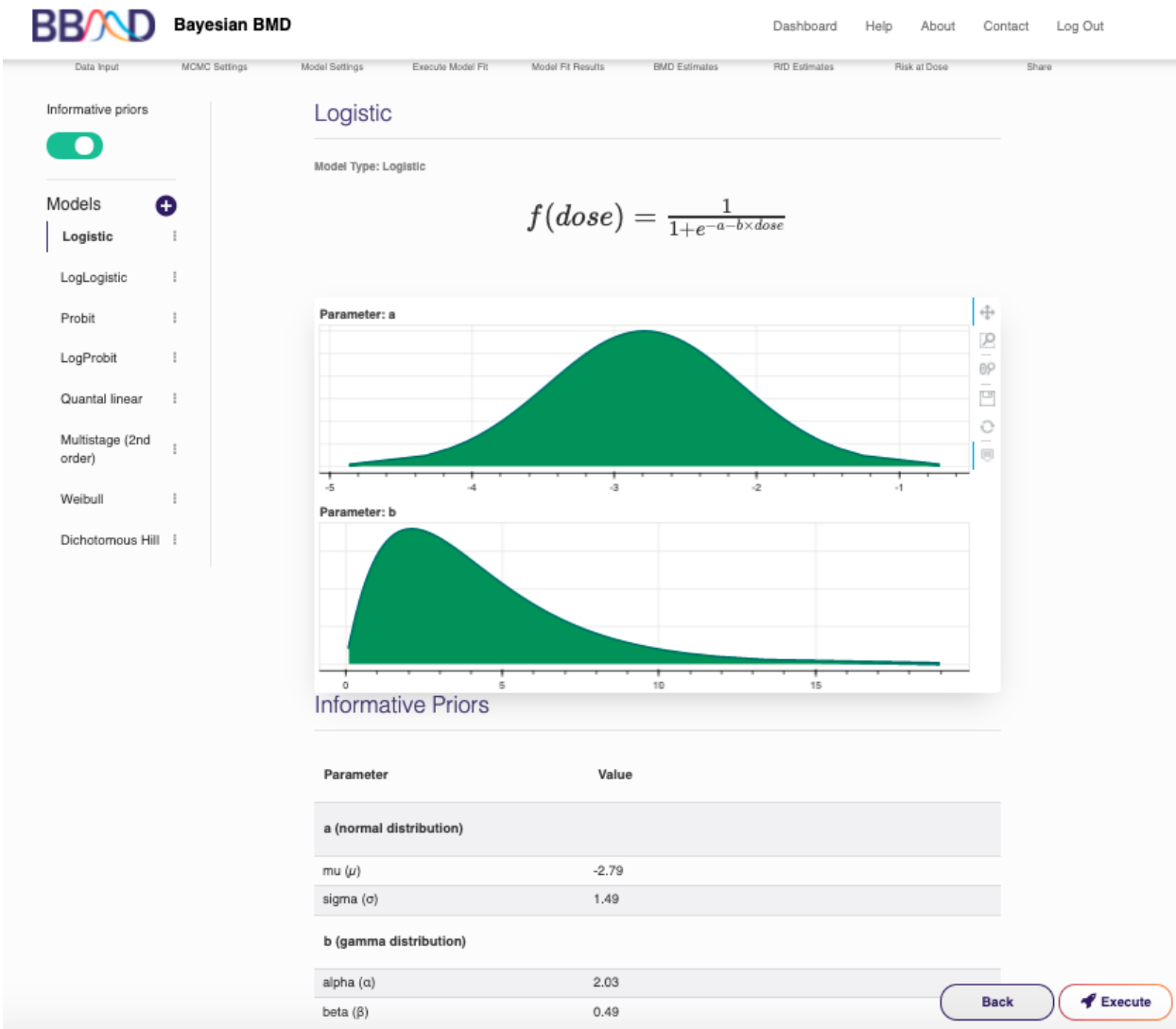


Figure 15. Distribution curves for dose-response model's informative priors

When you keep the default setting of the slider in the left column below "Informative Priors" as red, non-informative prior distribution is used as the default distribution for all the dose-response model parameters. The prior distributions for each model are listed below:

i. Dose-Response Models for Dichotomous Data

For Dichotomous data, there are eight models.

1) Quantal Linear Model:

$$f(dose) = a + (1-a) \times \left(1 - e^{-b\times dose}\right)$$
$$a \sim Uniform(0,1);\ b \sim Uniform(0,100)$$

2) Probit Model:

$$f(dose) = \Phi(a + b \times dose)$$
$$a \sim Uniform(-50, 50);\ b \sim Uniform(0, 100)$$

3) Logistic Model:

$$f(dose) = \frac{1}{1 + e^{(-a-b\times dose)}}$$
$$a \sim Uniform(-50, 50);\ b \sim Uniform(0, 100)$$

4) Weibull Model:

$$f(dose) = a + (1 - a) \times \left(1 - e^{-c\times dose^{b}}\right)$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(0, 50);\ b \sim Uniform(restriction, 15)$$

Where 'restriction' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

5) Multistage (2nd Order) Model:

$$f(dose) = a + (1 - a) \times \left(1 - e^{-b\times dose - c\times dose^{2}}\right)$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(0, 100);\ c \sim Uniform(0, 100)$$

6) LogLogistic Model:

$$f(dose) = a + \frac{1 - a}{1 + e^{-c-b\times log(dose)}}$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(restriction, 15);\ c \sim Uniform(-5, 15)$$

7) LogProbit Model:

$$f(dose) = a + (1 - a) \times \Phi(c + b \times log(dose))$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(restriction, 15);\ c \sim Uniform(-5, 15)$$

8) Dichotomous Hill Model:

$$f(dose) = a \times g + \frac{a - a \times g}{1 + e^{-c-b\times log(dose)}}$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(restriction, 15);\ c \sim Uniform(-5, 15);\ g \sim Uniform(0, 1)$$

These eight models are also the models which are part of the "Standard Models" for Dichotomous data.

ii. Dose-Response Models for Continuous Data

Background parameter 'a'

The background parameter "a" in all eight models has the same uniform distribution used as prior which is derived as follow:

The lower bound of the uniform distribution is always 0, and the upper bound is calculated differently for individual data and summary data.

For individual data,

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

i.e., doubling the largest response value in the input dataset.

For summary data,

$$a_{upper} = (\max(resp.mean) + 2 \times resp.sd_{mean.max}) \times 2$$

Where $\max(resp.mean)$ is the maximum mean response across all dose groups in the input dataset, and $resp.sd_{mean.max}$ is the response standard deviation in that dose group with the maximum mean response.

Slope parameter 'b'

For the Linear, Power, Michaelis Menten and Hill model, the lower or upper bound of the parameter b (a slope-equivalent parameter) are determined by the dose-response trend and the overall slope in the input data.

For individual input data and increasing trend:

$$b_{lower} = 0$$

$$b_{upper} = \frac{Max(resp) - Min(resp)}{Dose_{Max_resp} - Dose_{Min_resp}} \times 5$$

For individual input data and decreasing trend:

$$b_{lower} = \frac{Min(resp) - Max(resp)}{Dose_{Min_resp} - Dose_{Max_resp}} \times 5$$

$$b_{upper} = 0$$

Where $Max(resp)$ and $Min(resp)$ are the maximum and minimum response value in the input dataset. And $Dose_{Max_resp}$ and $Dose_{Min_resp}$ are the dose levels corresponding to the maximum and minimum responses respectively.

For summary input data:

$$b_{slope} = \frac{Mean_{Max_dose} + 2 \times SD_{Max_dose} - Mean_{Min_dose} - 2 \times SD_{Min_dose}}{Max(dose) - Min(dose)}$$

Where $Mean_{Max_dose}$ and SD_{Max_dose} are the mean and standard deviation of responses at the maximum dose level, and $Mean_{Min_dose}$ and SD_{Min_dose} are the mean and standard deviation of responses at the minimum dose level. $Max(dose)$ and $Min(dose)$ are the maximum and minimum dose levels in the input dataset. Because dose levels are first normalized to the scale between 0 and 1, $Max(dose)$ is very likely 1 and $Min(dose)$ is very likely 0. Then the prior distribution of the parameter "s" is $Uniform(0, 5 \times b_{slope})$ for increasing trend and $Uniform(5 \times b_{slope}, 0)$ for decreasing trend.

For continuous data, there are eight models.

1) Linear Model:

$$f(dose) = a + b \times dose$$
$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper})$$

2) Power Model:

$$f(dose) = a + b \times dose^{g}$$
$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper});\ g \sim Uniform(restriction, 15)$$

Where '**restriction**' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

3) Michaelis-Menten Model:

$$f(dose) = a + \frac{b \times dose}{c + dose}$$
$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper});\ c \sim Uniform(0, 15)$$

4) Hill Model:

$$f(dose) = a + \frac{b \times dose^{g}}{c^{g} + dose^{g}}$$
$$a \sim Uniform(0, a_{upper});\ b \sim Uniform(b_{lower}, b_{upper});\ c \sim Uniform(0, 15);\ g \sim Uniform(restriction, 15);$$

where '**restriction**' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

5) Exponential 2 Model:

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

$a \sim Uniform(0, a_{upper})$; $b \sim Uniform(0, 50)$ for **increasing trend** or $Uniform(-50, 0)$ for **decreasing trend**.

6) Exponential 3 Model:

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

$a \sim Uniform(0, a_{upper})$;
$b \sim Uniform(0, 50)$ for increasing trend or $Uniform(-50, 0)$ for decreasing trend;
$g \sim Uniform(restriction, 15)$;
where '**restriction**' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

7) Exponential 4 Model:

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

$a \sim Uniform(0, a_{upper})$; $b \sim Uniform(0, 100)$; $c \sim Uniform(0, 1)$ for decreasing trend or $c \sim Uniform(1, 15)$ for increasing trend.

8) Exponential 5 Model:

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

$a \sim Uniform(0, a_{upper})$; $b \sim Uniform(0, 100)$; $c \sim Uniform(0, 15)$ for decreasing trend or $c \sim Uniform(1, 15)$ for increasing trend; $g \sim Uniform(restriction, 15)$;
where '**restriction**' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

For continuous data, all eight of these models are included in the "Standard Models".

iii. Dose-Response Models for Categorical Data

1) Logistic Model:

$$f(dose) = \frac{1}{1 + e^{(-a - b \times dose)}}$$
$$a \sim Uniform(-50, 50);\ b \sim Uniform(0, 100)$$

2) Probit Model:

$$f(dose) = \Phi(a + b \times dose), b \geq 0$$
$$a \sim Uniform(-50, 50);\ b \sim Uniform(0, 50)$$

3) Cloglog Model:

$$f(dose) = 1 - e^{-e^{a+b\times dose}}$$
$$a \sim Uniform(-100, 100);\ b \sim Uniform(0, 100)$$

4) Quantal Linear Model:

$$f(dose) = a + (1-a) \times \left(1 - e^{-b\times dose}\right)$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(0, 100)$$

5) Multistage (2nd Order) Model:

$$f(dose) = a + (1-a) \times \left(1 - e^{-b\times dose - c\times dose^2}\right)$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(0, 100);\ \ c \sim Uniform(0, 100)$$

6) Weibull Model:

$$f(dose) = a + (1-a) \times \left(1 - e^{-c\times dose^b}\right)$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(restriction, 15);\ \ c \sim Uniform(0, 50);$$
where '**restriction**' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

7) LogLogistic Model:

$$f(dose) = a + \frac{1-a}{1 + e^{-c-b\times log(dose)}}$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(restriction, 15);\ \ c \sim Uniform(-5, 15);$$
where '**restriction**' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

8) LogProbit Model:

$$f(dose) = a + (1-a) \times \Phi(c + b \times log(dose))$$
$$a \sim Uniform(0, 1);\ b \sim Uniform(restriction, 15);\ \ c \sim Uniform(-5, 15);$$
where '**restriction**' is a user defined value and can be 0, 0.25, 0.5, 0.75 or 1.

These eight models are available to be used as dose-response models for categorical data, but the "Standard Models" are reduced to the Logistic, Probit, and Cloglog models.

The models shown on the left panel are the models will be analyzed by the system. Once you are happy with the models selected and the model settings, click "Execute" in the bottom right corner to execute model fitting.

E. Model Fit Results

On the “Model Fit Results” tab, the model fitting results obtained from the previous step are displayed. Click the name of one of the models on the left panel, then the results will be shown on the right (as shown in Figure 16) These results include the textual output of model parameter estimation, dynamic dose-response plot, posterior predictive p-value, model weight, correlation matrix, and graphical output of posterior sample of the model parameters (hidden by default). When click ‘Hide Parameters’, the parameter charts for each parameter in the model are displayed as shown in Figure 17.

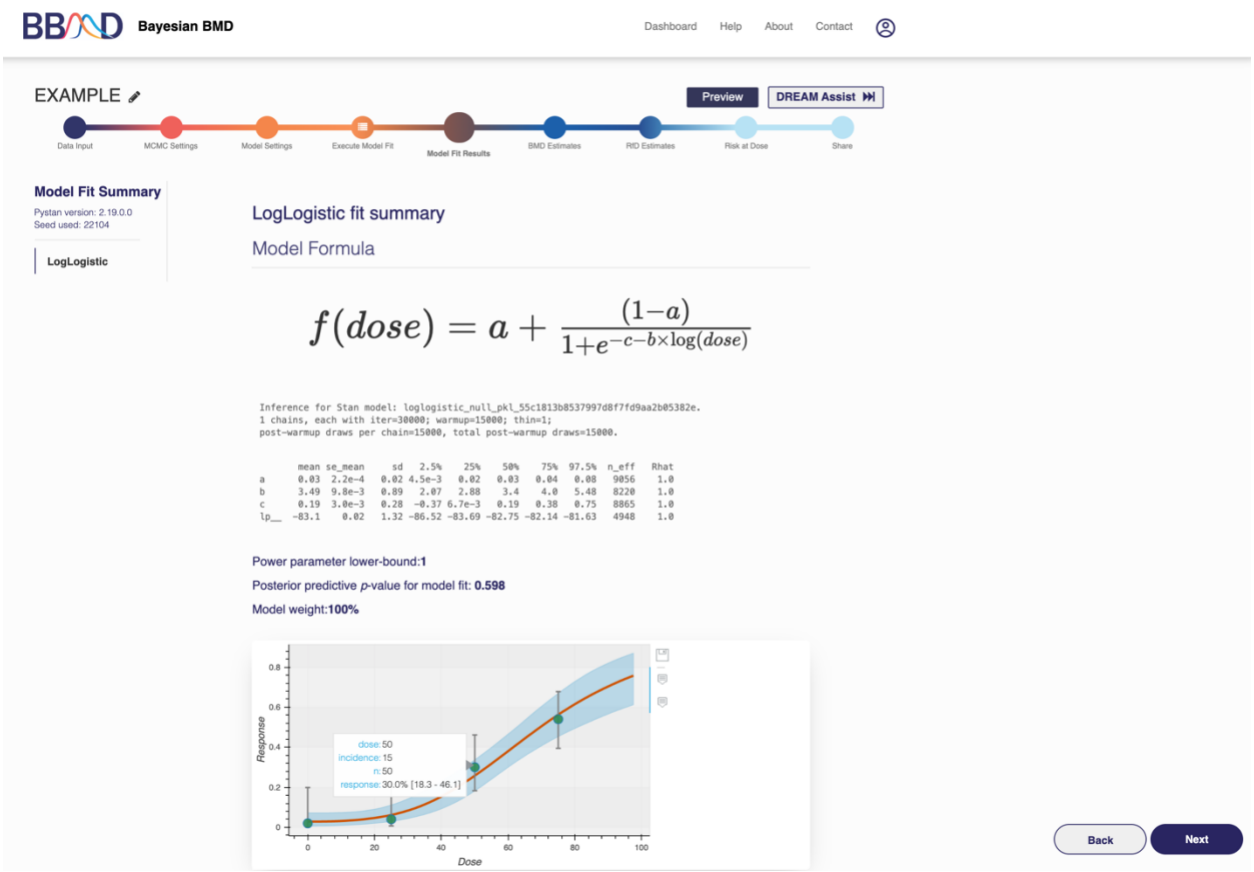


Figure 16. Results Shown on the “Model fit Results” Page

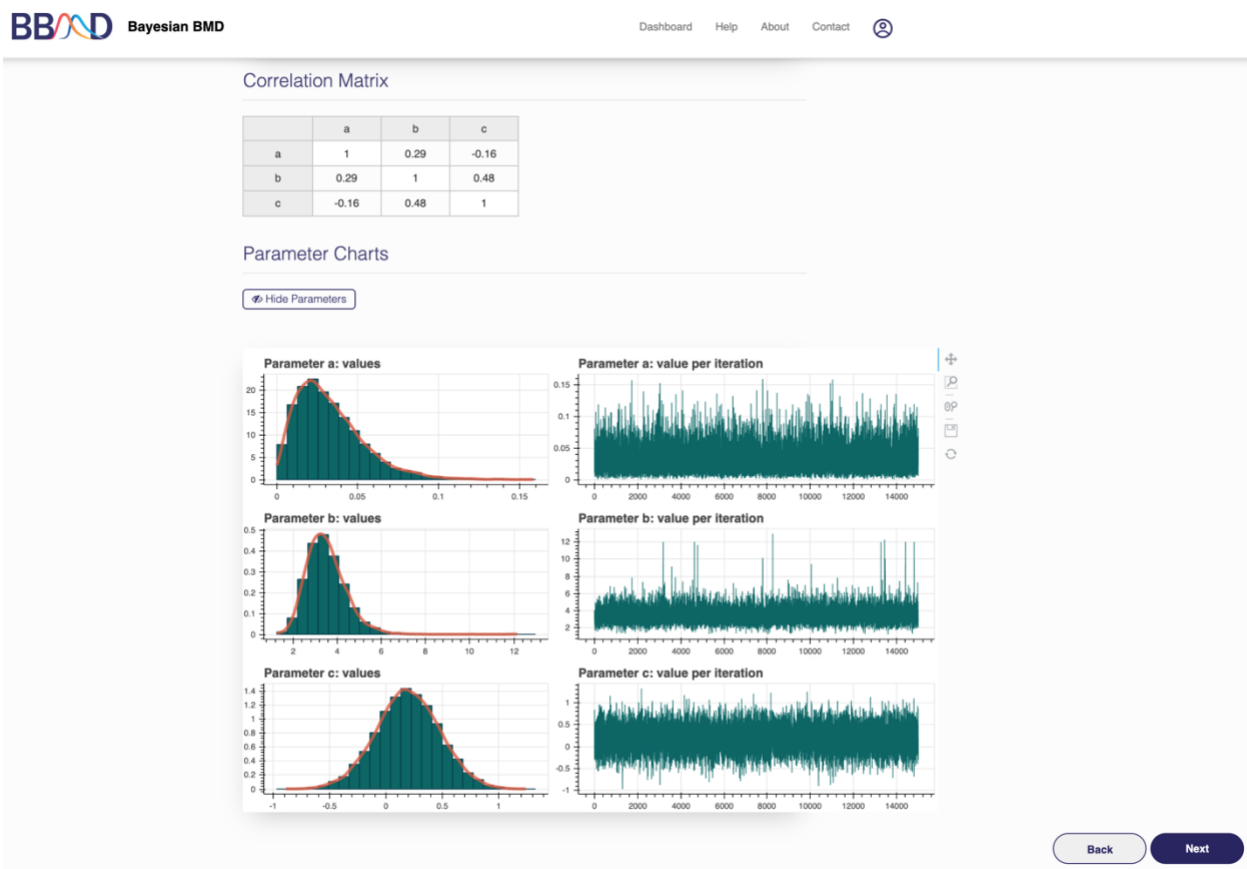


Figure 17. Parameter Charts

i. Parameter estimation results

The parameter estimation results, displayed in a table under the model formula, show the statistical summary for the estimated posterior distributions of parameters in the given dose-response model. These results are obtained direclty from PyStan's fit output, including some important statistics for model parameters and diagnostic indictors for the MCMC algorithms. The mean, standard error of the mean (se_mean), standard deviation (sd), various quantiles (2.5%, 25%, 50%, 75%, and 97.5%), and quantities indicating effective sample size (n_eff) and chain convergence (Rhat) for each model parameter derived from the posterior distribution of each parameter, as well as information regarding the MCMC execution are summarized in the table. As a note, the "Rhat" can be used to judge if the MCMC chains have converged properly. If the Rhat value is larger than 1.05, you may consider increasing the length of MCMC chains to get better convergence[1].

[1] 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

ii. Posterior Predictive P-Value

A posterior predictive p-value (PPP value) is reported below the dynamic dose-response plot. The PPP can be approximated by counting the predicted responses that satisfy the inequality out of the entire posterior sample space. This indicator can be used to judge if the fitting of this particular model is adequate. A large or small p-value means that a discrepancy in predicted data is very likely, further indicating a poor fit. Practically, if the PPP value is between 0.05and 0.95, then the fitting is adequate. The calculation procedure of PPP value 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 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 $\frac{M}{L}$

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

iii. Posterior Model Weight

A model weight ($\widehat{m}_j$) for model *j* is calculated for each model included in the analysis as a statistic for cross-model comparison. The model weight was introduced by (Wasserman, 2000), using the following two equations. The $\widehat{m}_j$ value of each selected model *j* is calculated as follows:

$$\widehat{m}_j = \exp\left(\widehat{\ell}_j - \frac{d_j}{2} log(n)\right),$$

where $\widehat{\ell}_j$ is a loglikelihood value estimated using one set of posterior samples of model parameters of the *j*th model, d_j is number of parameters in the *j*th model, and n is the sample size in the data set.

When all models in the analysis have an equal prior weight, the posterior model weight of model *j* is calculated by *m* value estimated from model *j* divided by the sum of *m* values estimated from all models in the analysis as the following equation.

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

This function assumes equal model priors for all models selected, so the weight mainly indicates how well the model fits the data. To make the weight more reliable, we use 1000 sets of randomly selected posterior samples of model parameters to calculate the model weights. This model weights are further applied to the model averaged BMD calculation in the F. BMD Estimation section.

iv. Interactive Dose-Response Plot

A dynamic dose–response plot is shown below the text box. This plot includes original dose–response data and a fitted curve with its 90th percentile interval shaded in blue. When you move your mouse over the dose-response curve, the estimated median and the 5th and 95th percentiles at a particular dose level will display. When you move your mouse over a data point from your inputted dataset, the dose, N, incidence, and the response percentile will also be displayed. Other information displayed in this figure includes the PyStan version, the lower bound placed on the power parameter (if applicable), the posterior predictive p-value (PPP value) for model fit and model weight for cross-model comparison.

v. Correlation Matrix

The fourth item displayed is the correlation matrix for the different model parameters. The correlation matrix is to show the correlation coefficients between different model parameters and is calculated using posterior samples.

vi. Plots for parameter posterior sample

If you click the "Show parameter charts", two plots (posterior sample trace plot and estimated probability density plot) will be displayed for each of the parameters in this dose-response model.

Basically, this is the results display tab, meaning that you can only review the results, not give the system additional inputs to modify the results.

F. BMD Estimation

On this page, you can calculate the BMD estimates of your interest. The settings for BMD calculation are slightly different between the analysis for dichotomous data, continuous data and categorical data, therefore, they will be introduced separately.

i. Dichotomous Data

a. <u>How to execute a BMD Analysis</u>

Figure 18 is a screenshot of the "BMD estimates" tab for a dichotomous dataset. To create a BMD analysis, you need to follow the four steps below:

(1) Name the BMD analysis using an easily identifiable name in the "BMD setting name" box.
(2) Specify a BMR value in the "Benchmark response value" box.

(3) Give prior model weight to the models included in this analysis. For your reference, 'Additional Info' section provides you further info on the performances of each model fitting. The prior weight, PPP value and the calculated weights for each model are provided when you change the slider in the right side of the page below "Additional Info" to green. The prior weight will influence the estimation of model averaged BMD value. Giving 0 prior weight to a particular model can exclude the model from model-averaged BMD calculation. The sum of the weights assigned to the individual models are not necessarily required to be 1. The system will automatically convert them. For example, if you give 1 to Logistic model and 3 to Probit model,the system will convert these values to 25% prior weight for the Logistic model and 75% prior weight for the Probit model.

(4) Click the "Execute" button to execute the BMD analysis using the settings just specified.

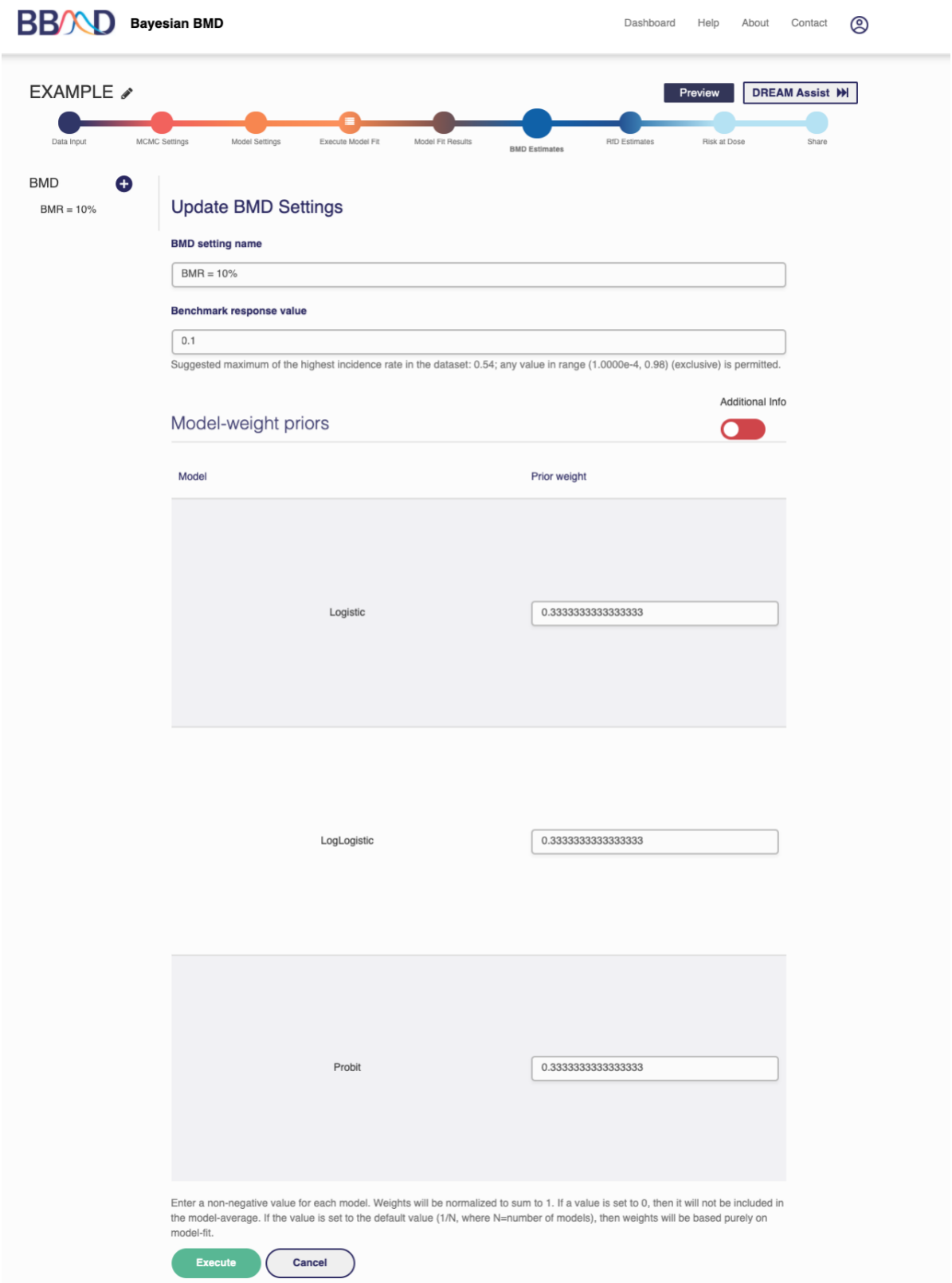


Figure 18. BMD Analysis Input for Dichotomous Data

Once the BMD analysis is successfully created, the name of the analysis will show on the left panel and the results will be displayed on the right panel, as shown in Figure 19. The results for a BMD analysis will include all the BMD distribution plots from each model, and a summary table. The estimation plots are shown in Figure 19, and the summary table is shown in Figure 20. To add a new BMD analysis, click the plus button in the left column and repeat steps (1) to (4) above. To edit or delete an existing BMD analysis, click the pencil in the upper right corner. From here you can change the name, BMR value, or model weights the same way as when the analysis was created. To save the changes made, click "Execute" on the bottom of the page. To delete this analysis, click "Delete". To cancel the modification, click 'Cancel'.

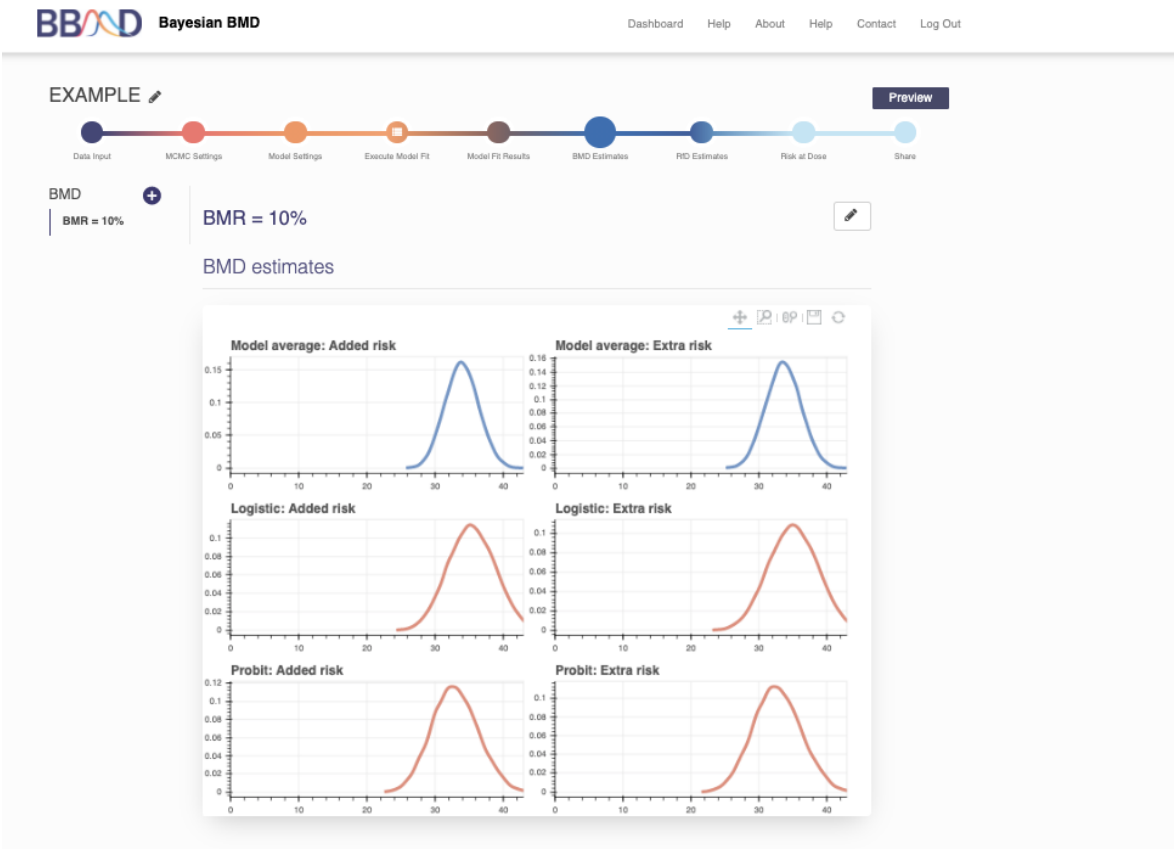


Figure 19. BMD Estimation plots from "BMD Estimates" tab

BMD summary table

Added risk

Statistic	Model average	Logistic	Probit
Prior model weight	N/A	0.50	0.50
Posterior model weight	N/A	0.40	0.60
Fraction with BMDs	100%	100%	100%
BMD (median)	**33.89**	**35.34**	**32.86**
BMDL (5th percentile)	**29.92**	**29.83**	**27.51**
25th percentile	32.24	33.02	30.57
Mean (SD)	33.92 (2.45)	35.40 (3.44)	32.94 (3.36)
75th percentile	35.56	37.75	35.19
95th percentile	38.02	41.12	38.66

Extra risk

Statistic	Model average	Logistic	Probit
Prior model weight	N/A	0.50	0.50
Posterior model weight	N/A	0.40	0.60
Fraction with BMDs	100%	100%	100%
BMD (median)	**33.62**	**35.03**	**32.63**
BMDL (5th percentile)	**29.45**	**29.19**	**27.03**
25th percentile	31.88	32.56	30.27
Mean (SD)	33.63 (2.55)	35.06 (3.60)	32.69 (3.49)
75th percentile	35.37	37.53	35.04
95th percentile	37.87	41.00	38.58

Back Next

Figure 20. BMD Estimation Results Shown on the "BMD Estimates" tab

b. Explanation of the analysis calculation

For dichotomous data, the BMD will be calculated for both the added risk and extra risk. For the two risks, the BMDs are defined by the following equations, respectively:

Added risk: $f(BMD) - f(0) = BMR$, Extra risk: $\frac{f(BMD)-f(0)}{1-f(0)} = BMR$;

where $f(.)$ represents a dichotomous dose–response model. BMR stands for benchmark response, which is a specified increase in the probability of response and is commonly set at 10%, 5%, or 1%.

In BBMD, the posterior distribution of BMD estimation is established. With the posterior sample, a number of statistics (including the mean, median, standard deviation, and other quantiles) of BMD can be computed and are reported in the 'BMD summary table' on the "BMD estimates"

tab. Based on our testing, the median value of the BMD posterior sample is the most reliable estimate for BMD owing to its resistance to some extreme values in the sample. In addition, the 5th percentile of the posterior sample is considered the lower bound of BMD (i.e., BMDL) corresponding to the lower bound of the one-sided 95th confidence interval. The BMDL is usually used as the point of departure for low-dose extrapolation and is therefore of great regulatory interest. The BMD and BMDL values are highlighted in the 'BMD summary table'. The same procedures used for determining BMD and BMDL are also applied to continuous and categorical data.

ii. Continuous Data

a. How to execute a BMD Analysis

For continuous data, the BBMD system provides two ways to define BMD: (1) based on central tendency and (2) based on tails (hybrid approach). The basic steps to create a BMD analysis for continuous data are almost identical to the procedure for dichotomous data, except the settings for the BMR. "Central tendency" is the default option for BMD estimation method of continuous data.

Figure 21 is a screenshot of the "BMD estimates" tab for a continuous dataset using Central tendency as the BMD estimation method. To create a BMD analysis, you need to follow the six steps below:

(1) Name the BMD analysis using an easily identifiable name in the "BMD setting name" box.
(2) Specify "Central Tendency" as the estimation method to be used in the "BMD estimation method" drop-down menu. If you wish to use "Hybrid Method (tails)", the detailed steps are in the next section.
(3) Select an adversity measure from the "Adversity measure" drop-down menu. For BMD defined on central tendency, there are three options for defining the BMR value: a) relative change, b) absolute change, and c) cutoff. Detailed explanations on the three options are in section b.
(4) Specify an adversity value in the "Adversity value" text box.
(5) Give prior model weight to the models included in this analysis. 'Additional Info' section provides you further info on the performances of each model fitting. The prior weight, PPP value and the calculated weights for each model are provided when you change the slider in the right side of the page below "Additional Info" to green. The prior weight will influence the estimation of model averaged BMD value. Giving 0 weight to a particular model can exclude the model from model-averaged BMD calculation. The sum of the weights assigned to the individual models are not necessarily required to be 1. The system will automatically convert them. For example, if you give 1 to Linear model and 3 to Exponential 2 model,the system will convert these values to 25% prior weight for the Linear model and 75% prior weight for the Exponential 2 model.
(6) Click the "Execute" button to execute the BMD analysis using the settings just specified.

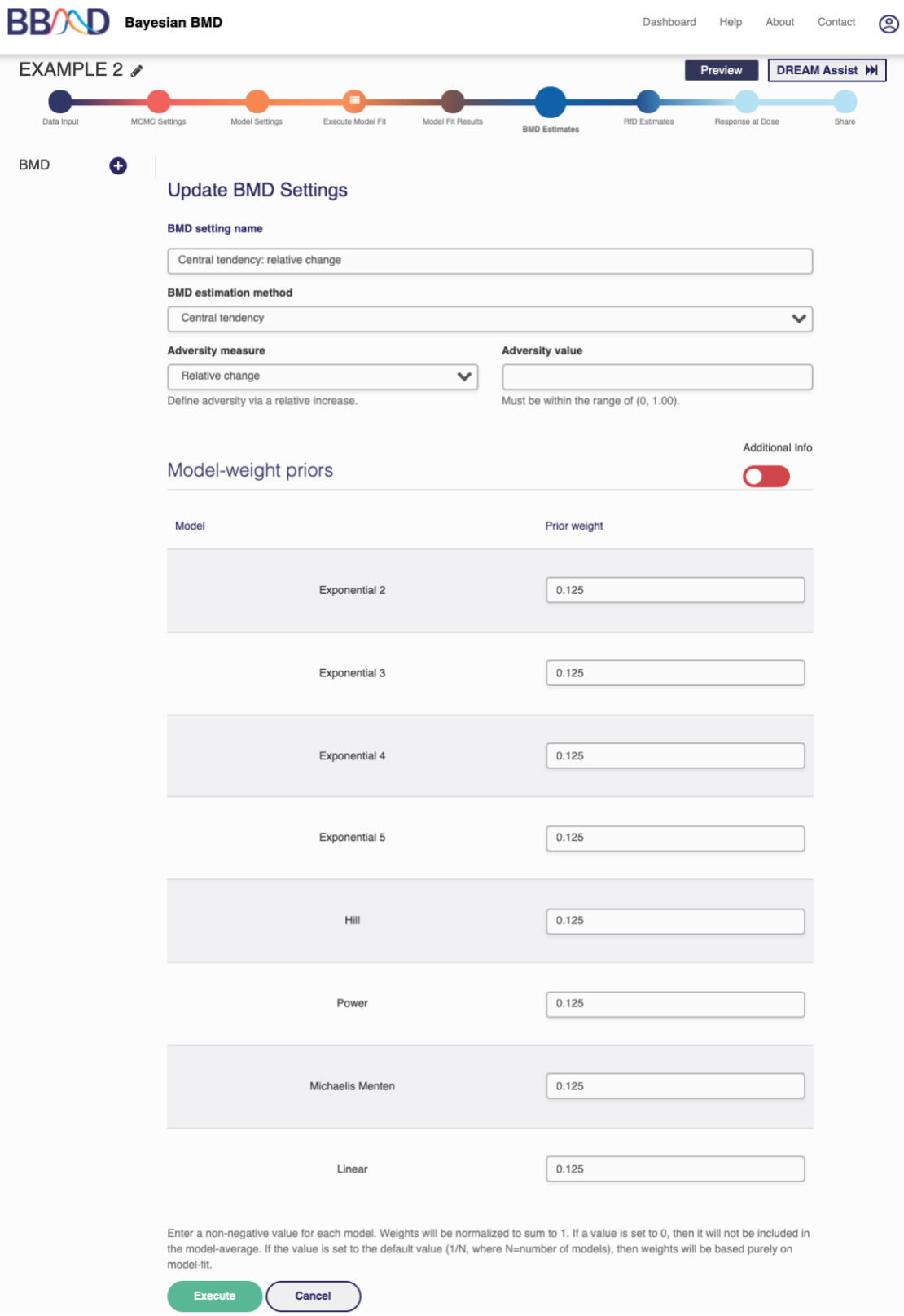


Figure 21. Settings for Continuous Data BMD Calculation Based on Central Tendency

Figure 22 is a screenshot of the “BMD Estimates” tab for a continuous dataset using the hybrid method (tails) as the estimation method. To create a BMD analysis, follow the seven steps below:

(1) Name the BMD analysis using an easily identifiable name in the “BMD setting name” box.
(2) Specify “Hybrid Method (tails)” as the estimation method to be used in the “BMD estimation method” drop-down menu.
(3) Select an adversity measure from the “Adversity measure” drop-down menu. For BMD defined on hybrid method, there are two options for defining the BMR value: a) control group percentile, and b) absolute cutoff value.
(4) Specify an adversity value in the “Adversity value” text box.
(5) Specify a BMR value in the “Benchmark response value” text box.
(6) Give prior model weight to the models included in this analysis. ‘Additional Info’ section provides you further info on the performances of each model fitting. The prior weight, PPP value and the calculated weights for each model are provided when you change the slider in the right side of the page below “Additional Info” to green. The prior weight will influence the estimation of model averaged BMD value. Giving 0 weight to a particular model can exclude the model from model-averaged BMD calculation. The sum of the weights assigned to the individual models are not necessarily required to be 1. The system will automatically convert them. For example, if you give 1 to Linear model and 3 to Exponential 2 model,the system will convert these values to 25% prior weight for the Linear model and 75% prior weight for the Exponential 2 model.
(7) Click the “Execute” button to execute the BMD analysis using the settings just specified.

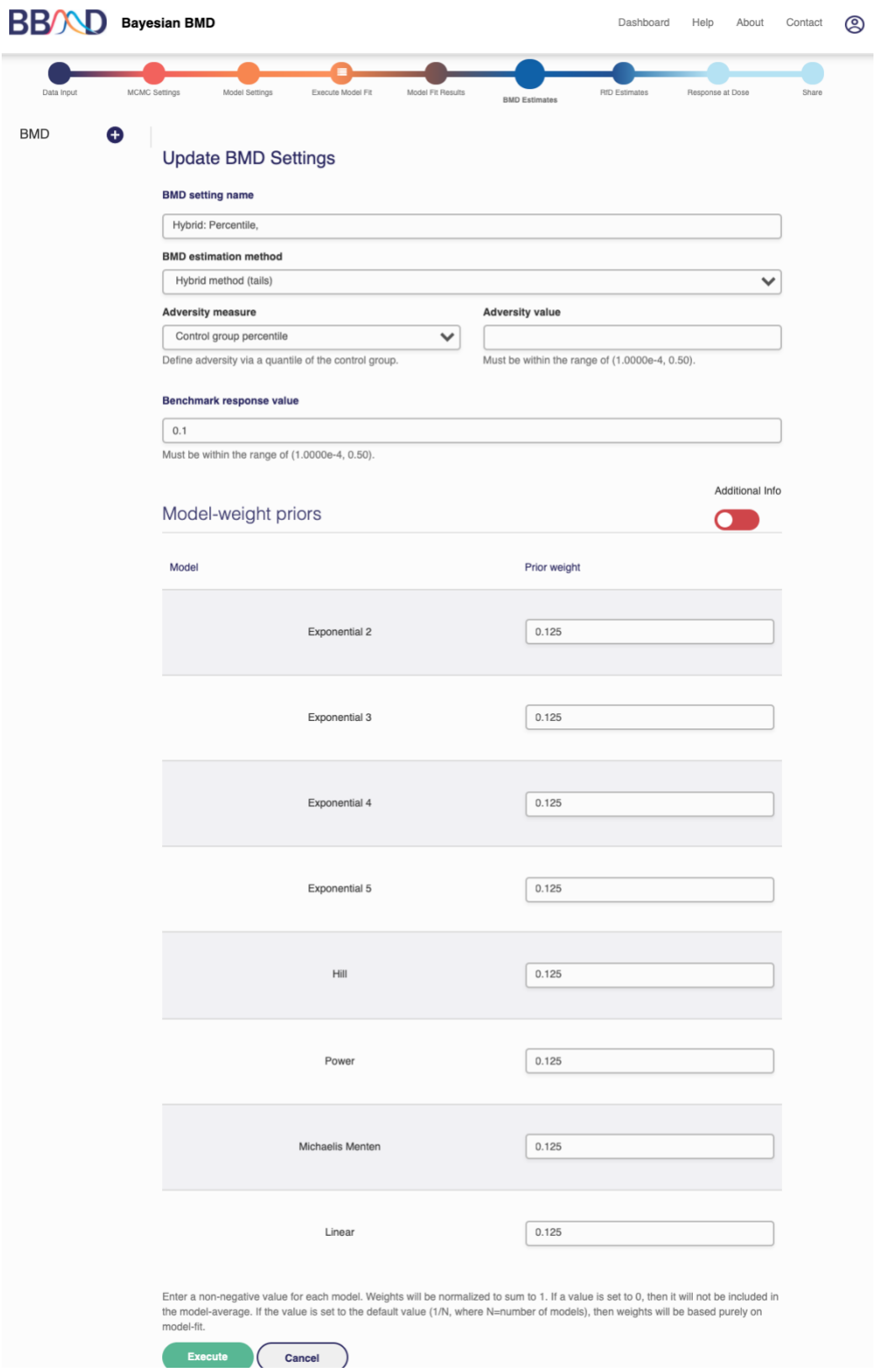


Figure 22. Settings for Continuous Data BMD Calculation Based on Tails

Once the BMD analysis is successfully created, the name of the analysis will show on the left panel and the results will be displayed on the right panel (shown previously in Figure 19). To add a new BMD analysis, click the plus button in the left column and repeat steps (1) to (6) above. To edit or delete an existing BMD analysis, click the pencil in the upper right corner. From here you can change the name, estimation method, adversity measure, adversity value, bmr value, or model weights the same way as when the analysis was created. To save the changes made, click "Execute" on the bottom of the page. To delete this analysis, click "Delete". To cancel the modification, click 'Cancel'.

b. Explanation of the analysis calculation

For both types of estimation methods, an adversity measure needs to be specified. For "Central Tendency" there are three options to specify an adversity:

1) Relative change:
   For this option, you 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:

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

where $f(.)$ represents a continuous dose–response model fit to the central tendency of the data (i.e., the median under the lognormality assumption). The BMD stands for the dose level that satisfies the selected definition equation. 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) Absolute change:
   For this option, you need to input a value of absolute change, e.g., 3.2. This means that if the centraltendency 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 itis "+", otherwise it is "-".

3) Cutoff:
   For this option, you 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 becalculated accordingly, using the following equation:

$$f(BMD) = cutoff$$

   The allowable range for the values of these three options will be automatically calculated based onthe trend of the dose-response data and shown.

For the "Hybrid Method (tails)" estimation method, an adversity value must be specified in addition to a BMR value. The hybrid approach considers any response above or below (i.e., corresponding to increasing or decreasing trend) the adversity value as abnormal; thus, the BMD is the dose level where the proportion of the abnormality has increased a certain percent (i.e., BMR) compared with the control. There are two options to specify an adversity:

1) Absolute cutoff value:
   For this option, you need to input a value of a cutoff. Then, depending on increasing or decreasing dose-response trend, above or below this value will be considered as adverse.

2) Control group percentile:
   For this option, you need to input a percentile value of the control. Then, the below 1st percentile or above 99th percentile of the control distribution is considered as adverse depending on decreasing or increasing.

Mathematically, for increasing trend, the hybrid BMD definition can be expressed as

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

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

For decreasing trend,

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

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

where $Q(0)$ represents the quantile of the adversity value at control dose and $Q(BMD)$ represents the quantile of the adversity value at the BMD level.

iii. Categorical Data

a. How to execute a BMD Analysis

Figure 23 is a screenshot of the "BMD estimates" tab for a categorical dataset. To create a BMD analysis, you need to follow the five steps below:

(1) Name the BMD analysis using an easily identifiable name in the "Model name" box.
(2) Specify a BMR value in the "Benchmark response value" box.
(3) Select a severity level from the "Severity Level" drop-down list. The number of options equals to levels of the input dataset.
(4) Give prior model weight to the models included in this analysis. Similarly, 'Additional Info' section provides you further info on the performances of each model fitting. The prior

weight, PPP value and the calculated weights for each model are provided when you change the slider in the right side of the page below "Additional Info" to green. The prior weight will influence the estimation of model averaged BMD value. Giving 0 weight to a particular model can exclude the model from model-averaged BMD calculation. The sum of the weights assigned to the individual models are not necessarily required to be 1. The system will automatically convert them. For example, if you give 1 to Logistic model and 3 to Cloglog model,the system will convert these values to 25% prior weight for the Logistic model and 75% prior weight for the Cloglog model.

(5) Click the "Execute" button to execute the BMD analysis using the settings just specified.

Once the BMD analysis is successfully created, the name of the analysis will show on the left panel and the results will be displayed on the right panel (previously shown in Figure 20). To add a new BMD analysis, click the plus button in the left column and repeat steps (1) to (5) above. To edit or delete an existing BMD analysis, click the pencil in the upper right corner. From here you can change the name, BMR value, severity level, or model weights the same way as when the analysis was created. To save the changes made, click "Execute" on the bottom of the page. To delete this analysis, click "Delete". To cancel the modification, click 'Cancel'.

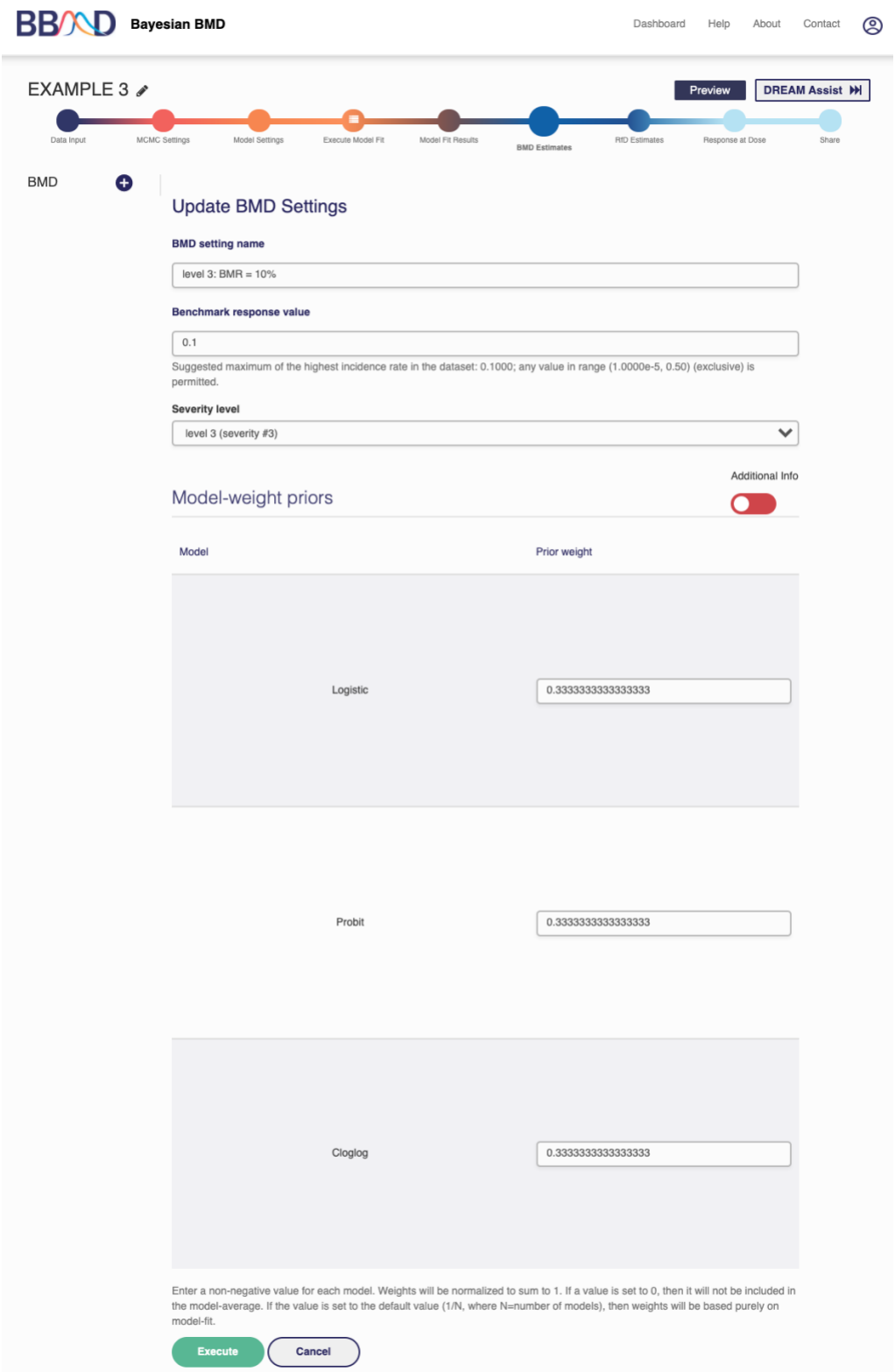


Figure 23. Settings for a Categorical Data BMD

b. Explanation of the analysis calculation

The BMD is defined based on "Central Tendency" for each severity level. The BMD are calculated for both the added risk and extra risk at each severity level. For the two risks, the BMDs are defined by the following equations, respectively:

Added risk: $f(BMD) - f(0) = BMR$, Extra risk: $\frac{f(BMD)-f(0)}{1-f(0)} = BMR$;

where $f(.)$ represents a dichotomous dose–response model. BMR stands for benchmark response, which is a specified increase in the probability of response and is commonly set at 10%, 5%, or 1%.

iv. Model Averaged BMD Calculation

Before calculating a model averaged BMD, a posterior sample of the BMD from each individual model should be obtained. The process to get the posterior sample was described in the previous three sections. Here, we focus on the method to get model averaged BMD.

a. Posterior Model Weight for Model Averaged BMD 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 the previous section.

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

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. That is, the model with 0 weight will be excluded from the analysis.

b. Model Averaged BMD Calculation

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

$BMD_{1-1}, BMD_{1-2}, \dots, BMD_{1-15000}$ for model 1
$BMD_{2-1}, BMD_{2-2}, \dots, BMD_{2-15000}$ for model 2
…

Then the posterior sample of model averaged BMD is calculated as a mixture distribution over all models:

$$\Pr(BMD_{ma}|Data) = \sum_{j=1}^{J} \Pr(BMD_j|M_j, Data)\Pr(M_j|Data),$$

That is:

$$BMD_{MA-15000} = BMD_{1-15000} \times w_1 + BMD_{2-15000} \times w_2 + \cdots$$

Therefore, we will have the same size of posterior sample for model averaged BMD. $w_1, w_2, \ldots$ are posterior model weight (prior model has been integrated) calculated in the previous section.

G. Probabilistic RfD Estimation

On this page, you can calculate the RfD estimates of your interest using your previously calculated BMD estimates. The RfD estimate settings are the same for each data type.

Figure 24 is a screenshot of the "RfD estimates" tab. To create a RfD analysis, you need to follow the eleven steps below:

1) Name the RfD analysis in the "RfD analysis name" box
2) Specify the dose units in the "Dose Units" box
3) (Optional) Choose a random number seed for the "Random number seed" box
4) Select a BMD analysis to be used as the point of departure. Your BMD analyses from the previous tab will be the possible selections in the drop-down menu.
5) Specify the Allometric Scaling settings. Select a test species from the "Test species" drop-down menu. Specify the test species body weight in the "Test species body weight" box. Choose a human body weight in the "Human body weight" box. Lastly, specify the Allometric scaling exponent mean and standard deviation in the two corresponding boxes. For each species you choose the default values for the parameters in 'Allometric Scaling' section will automatically be displayed.
6) Give an Animal to Human Uncertainty geometric mean and geometric standard deviation. These two values go in their respective boxes.
7) Specify Duration of Exposure details to extrapolate non-chronic exposures to chronic exposures. Select the duration of exposure from the drop-down menu. Then give the short-term exposure geometric mean and geometric standard deviation in the corresponding boxes.
8) (Optional) Add up to two additional uncertainties. Each additional uncertainty requires a name, geometric mean, and geometric standard deviation. These choices go in the respective boxes.
9) Specify the Human Variability geometric mean and geometric standard deviation in the corresponding boxes.
10) Give the target population-based incidence in the I* box.
11) Click the "Save" button to execute the RfD analysis using the settings just specified.

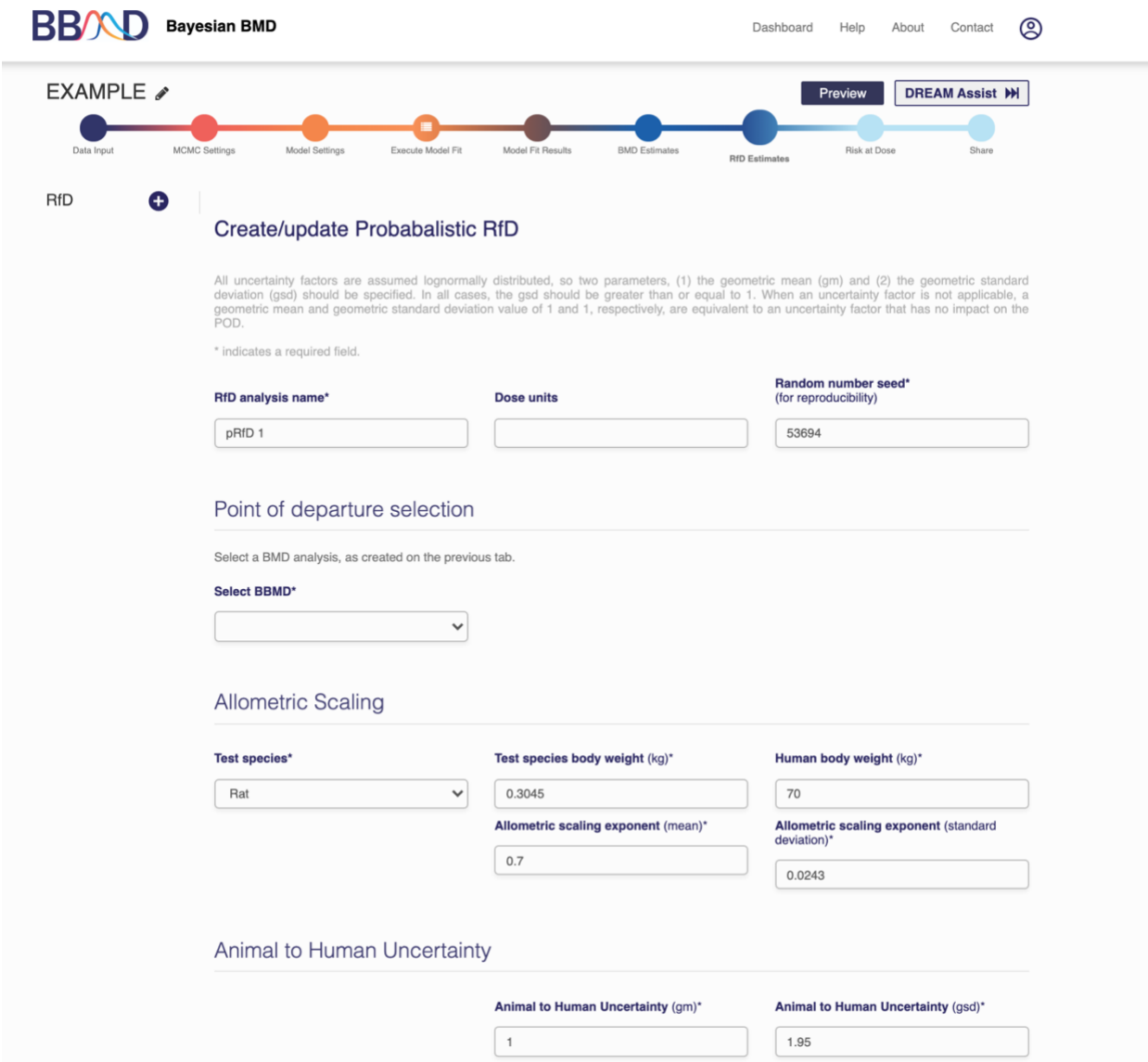


Figure 24. Beginning of the RfD analysis tab

Once the RfD analysis is successfully created, the name of the analysis will show on the left panel and the results will be shown on the right panel as shown in Figure 25. To add a new RfD analysis, click the plus button in the left column and repeat steps (1) to (11) above. To edit or delete an existing RfD analysis, click the pencil in the upper right corner. From here you can change all analysis settings the same way as when the analysis was created. To save the changes made, click “Update” on the bottom of the page. To delete this analysis, click “Delete”.

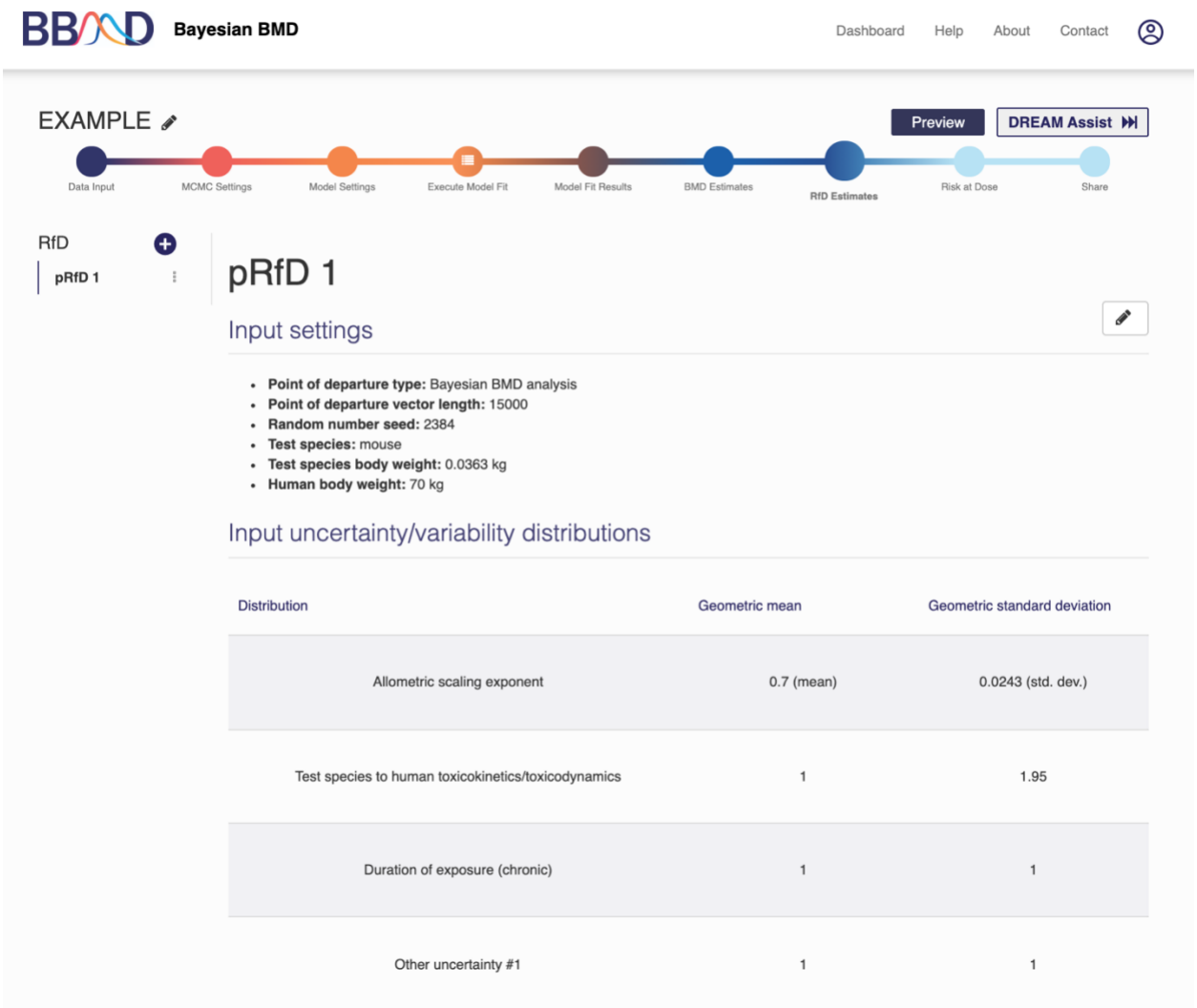


Figure 25. RfD Estimate results shown after RfD analysis finishes

H. Risk/Response at Dose

On this page, you can calculate the RAD (Response at Dose) estimates of your interest. The settings for RAD analyses using dichotomous data and continuous data are identical. The only difference from these two data types and categorical data is that a specified severity level is required for categorical data.

Figure 26 is a screenshot of the "Risk at Dose" tab (This tab is called "Response at Dose" for continuous and categorical data). To create a new RAD analysis, you need to follow the five steps below:

1) Name the RAD analysis using an identifiable name in the "Model Name" box.
2) Specify a dose value in the "Dose value" box.
3) Only for categorical data – Choose a severity level from the "Severity level" drop-down menu.

4) Give prior model weights to the models included in this analysis. Giving 0 weight to a particular model can exclude the model from model-averaged RAD calculation. The sum of the weights assigned to the individual models are not necessarily required to be 1. The system will automatically convert them.
5) Click the "Save" button to execute the RAD analysis using the settings just specified.

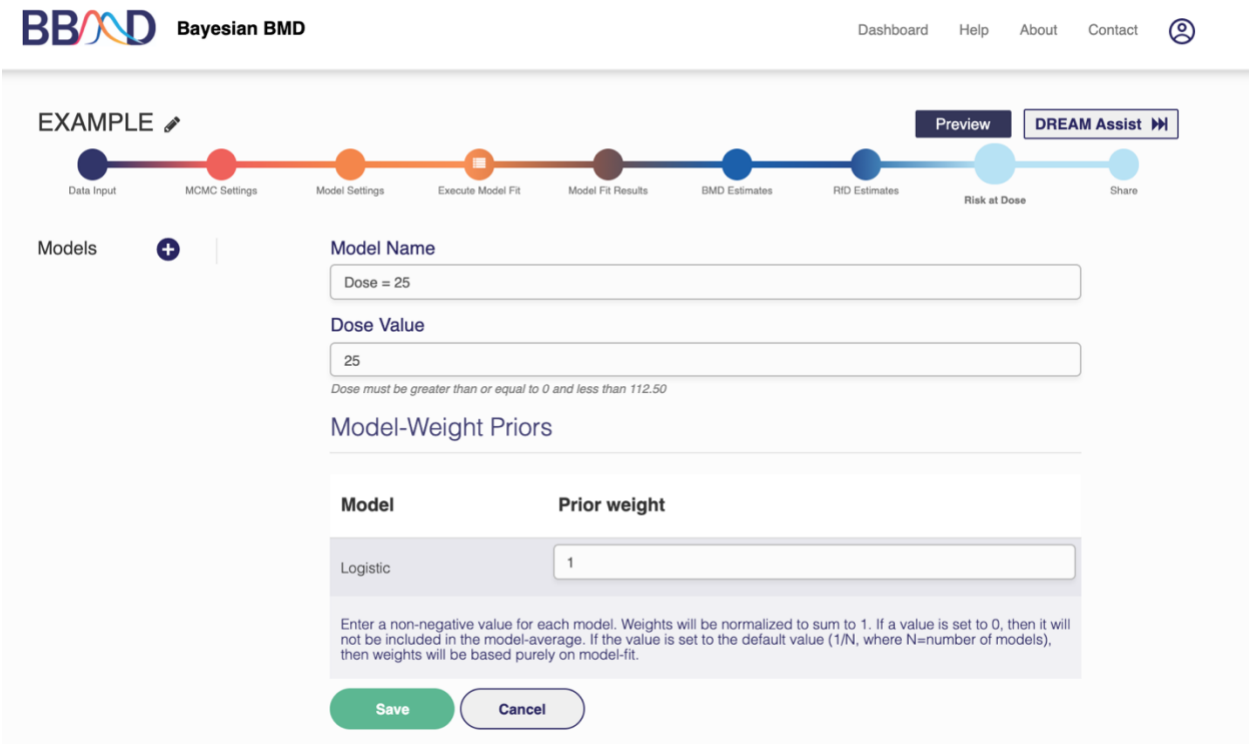


Figure 26. Risk at Dose analysis for Dichotomous Data

Once the RAD analysis is successfully created, the name of the analysis will be shown on the left panel and the results will be shown on the right as seen in Figure 27. To add a new RAD analysis, click the plus button in the left column and repeat steps (1) to (5) above. To edit or delete an existing RAD analysis, click the pencil in the upper right corner. From here you can change all analysis settings the same way as when the analysis was created. To save the changes made, press "Update" on the bottom of the page. To delete this analysis, press "delete".

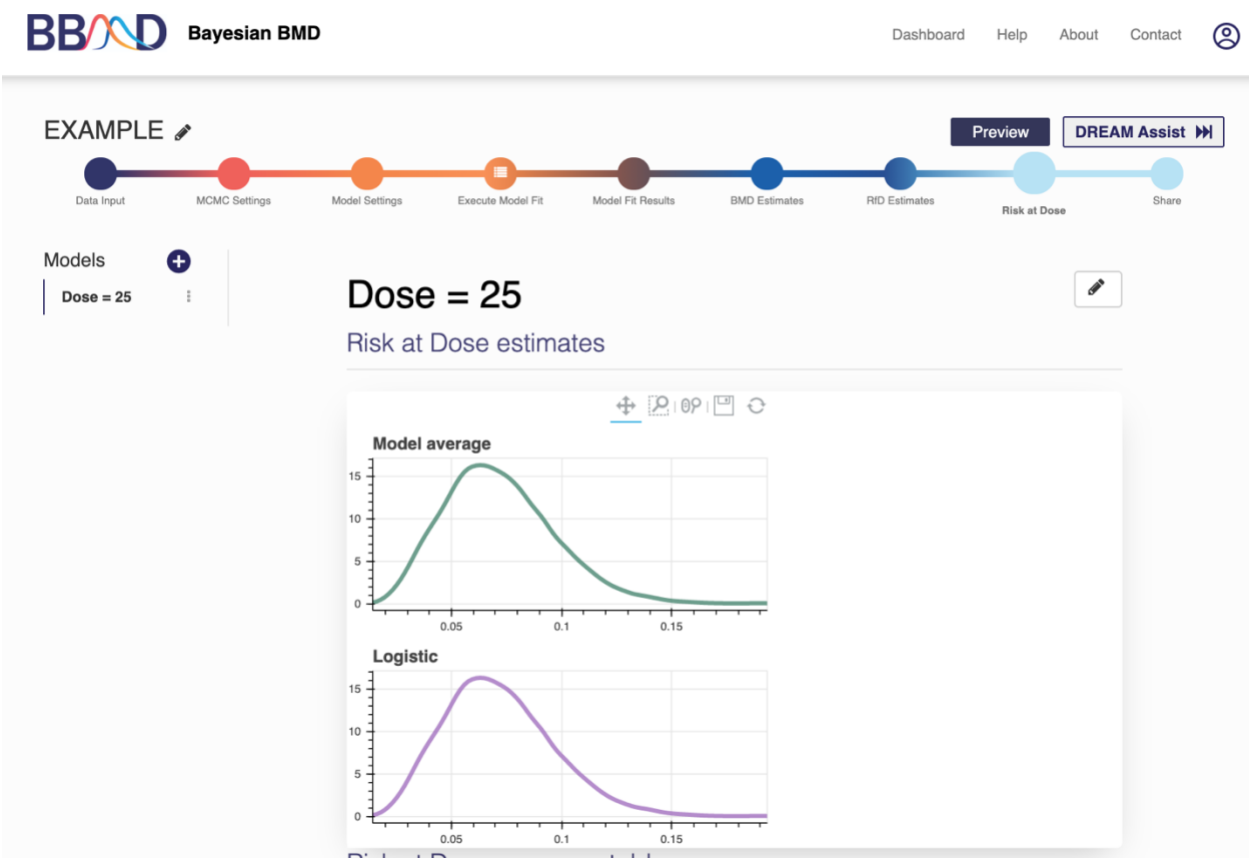


Figure 27. RAD Estimate results shown in on the "Risk at Dose" tab

I. Share the Analysis

The final tab for this analysis is the "Share" tab, shown in Figure 28. By default, all analyses in a personal account can only be accessed by the owner of the account. If you would like to share your analysis with others, you can change the settings from "Private" to "Public", "Organization", or "Dream Assist". The public setting allows you to send the created URL to others for them to access and review (but not edit) the analysis. The 'Organization' setting allows you to share the link with others for them to edit the analysis. Lastly, the DREAM setting will share this analysis with the DREAM Tech team, who can assist with the analysis. The 'Organization' setting is coming soon.

You can also export the results of the analysis into Word or Excel formats. Before exporting results, you can customize the parts of the analysis included on the reports. By clicking "show more" on the right side of the page, the customization options will appear. Here you can specify the BMD results, RAD results, or model summaries to be included. If you wish to change the report settings in the future, you can return to this analysis and export a new report. Exporting the results will send a link to your account's email where you can download the reports.

At any time during the updating or reviewing stage, if you want to change to another existing analysis, you can click the “Dashboard” button on the top right corner to switch to the summary page for the existing analyses and access another analysis.

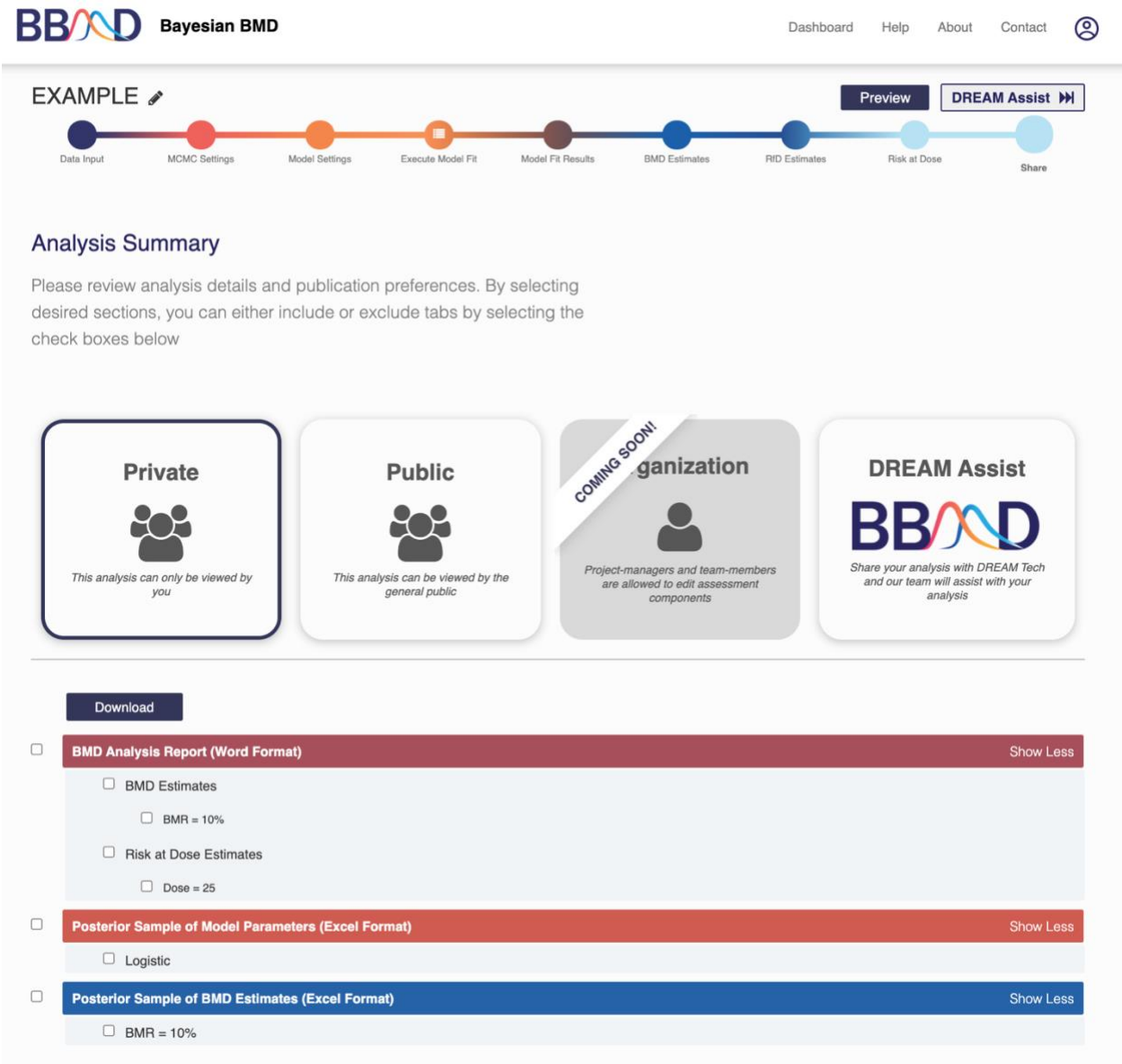


Figure 28. “Share” tab for a BMD analysis using a single dataset

III. Batch Processing for Analysis

An automatically generated name "New batch, *Month Day Year, HH:MM PM*", is assigned to the newly started analysis. You can click the pencil button next to the analysis name, as seen in Figure 29, to make the name more identifiable. Without inputting any data, you will not be able to advance through the tabs. Therefore, the first step in the analysis is to input the batch dose-response data.

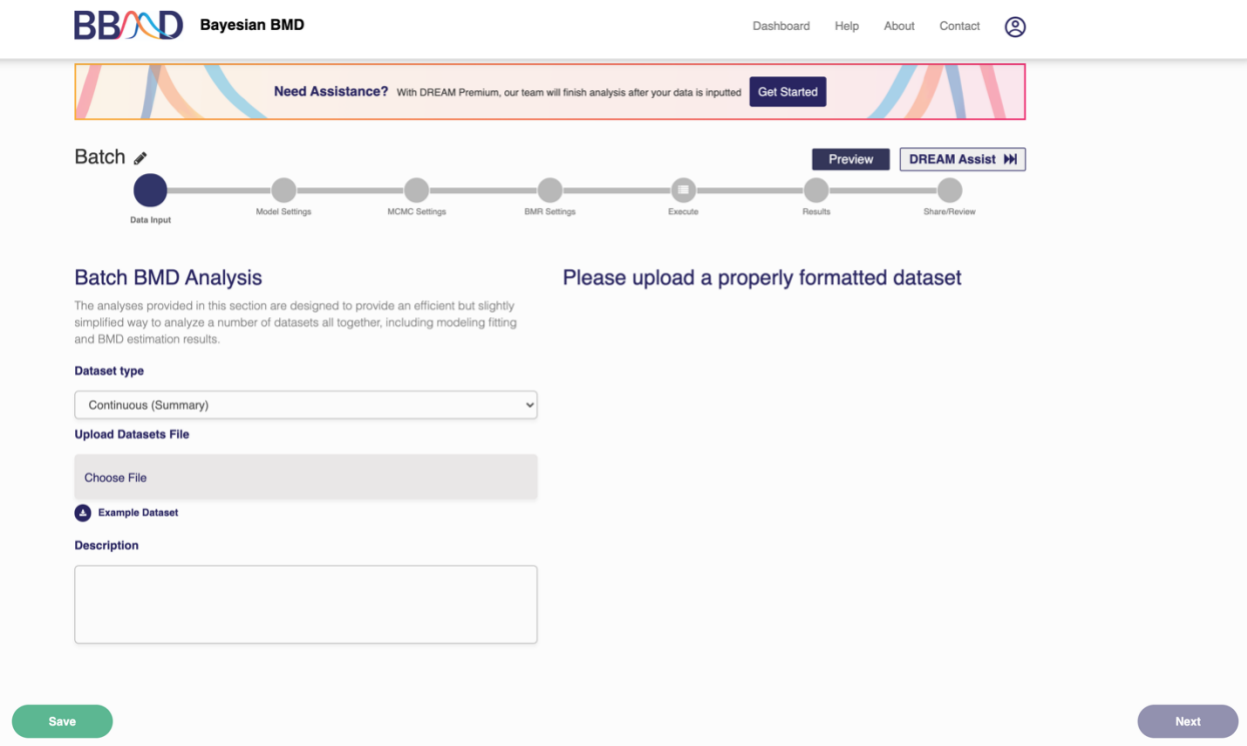


Figure 29. Start of a new Batch Analysis

A. Data Input

Note: For data input, a ".csv" file needs to be uploaded. You can click 'Example Dataset' to see the required format for each type of dataset.

If you choose to do an analysis on dichotomous data, only dichotomous summary data is allowed. Five columns are required for input (from left to right): Dataset, Index, Dose level, number of subjects, number of subjects affected.

If you choose to do an analysis on continuous data, both continuous summary and continuous individual data is accepted. For continuous summary data, six columns are required. The columns (left to right) are Dataset, Index, dose level, number of subjects, mean, standard deviation. For continuous individual data, four columns are required (from left to right): Dataset, Index, dose level, response.

If you choose to do an analysis on categorical data, five columns are required (left to right): Dataset, Index, dose level, severity level, response.

After inputting your dataset, you can add a description to the analysis in the “Description” box. When the dataset is successfully uploaded and a description is added, press “Next” in the lower right corner to advance to the next tab.

B. Model Settings

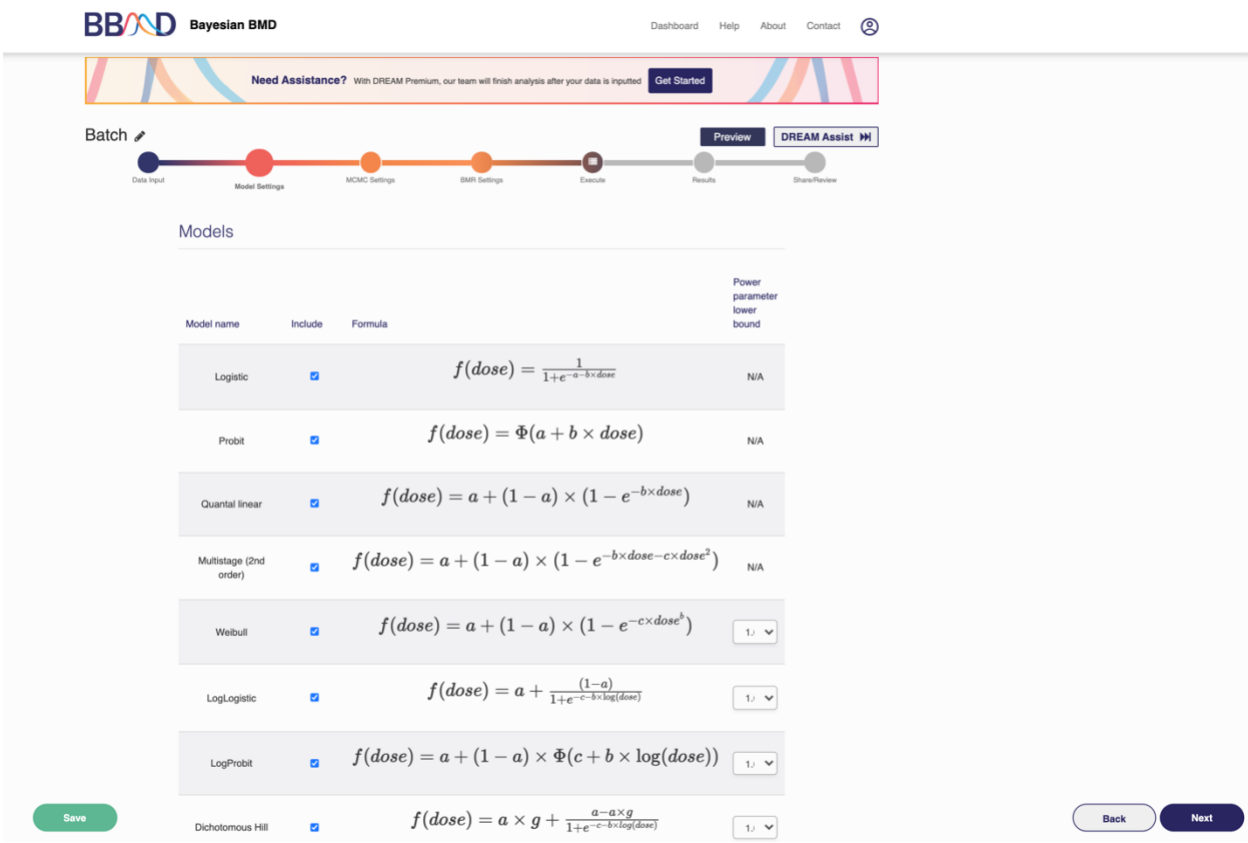


Figure 30. “Model Settings” tab

Figure 30 shows the “Model Settings” tab, the next step in performing a batch BMD analysis. To include a model in the analysis, ensure that the check box is checked. The model formula is in the next column. For models there is a power parameter lower bound, you can choose a value in the last column. The power parameter lower bound options are 0, 0.25, 0.5, 0.75, and 1.

The models available for each data type are slightly different. For dichotomous data the eight available models are: Logistic, Probit, Quantal Linear, Multistage (2nd Order), Weibull, LogLogistic, LogProbit, Dichotomous Hill. For continuous data the eight available models are: Linear, Power, Michaelis-Menten, Hill, Exponential 2, Exponential 3, Exponential 4, Exponential 5. For categorical data the eight available models are Logistic, Probit, Cloglog, Quantal Linear, Multistage (2nd order), Weibull, LogLogistic, LogProbit.

C. MCMC Settings

On this tab you can specify the settings for the MCMC algorithm.

“Iterations” is the length of MCMC chain, i.e., the number of posterior samples in each MCMC chain. Default value is 30,000. The allowable range is any integer between 10,000 and 50,000.

“Number of chains” is the number of Markov Chains to be sampled. Default value is 1. Allowable rangeis 1 - 3.

“Warmup percent (%)”, the percent of sample in each Markov Chain will be discarded from the final posterior sample. Default value is 50% with an allowable range of 10% - 90%.

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

$$30000 \times 1 \times (1 - 50\%) = 15000$$

“Seed” is random seed number used in the MCMC algorithms. The number is randomly generated, butyou can specify the number for the purpose of reproduction.

Once these values are specified, click “Save” to save the MCMC settings.

D. BMR Settings

The BMR settings tab, shown in Figure 31, is where you specify the BMR type and BMR values for the analysis. To add a new benchmark dose calculation, click the plus button on the right side of the right panel. When adding a new BMR, specify either Extra or Added for the benchmark dose calculation, then specify the BMR value. To delete a BMR, click the red trash can button on the right side of the table. You can add up to 4 different BMRs for the analysis.

When you are satisfied with the specified BMRs, either press “Save” in the bottom left corner of the page to save the analysis to come back to later, or press “Execute” in the bottom right corner to execute the batch analysis.

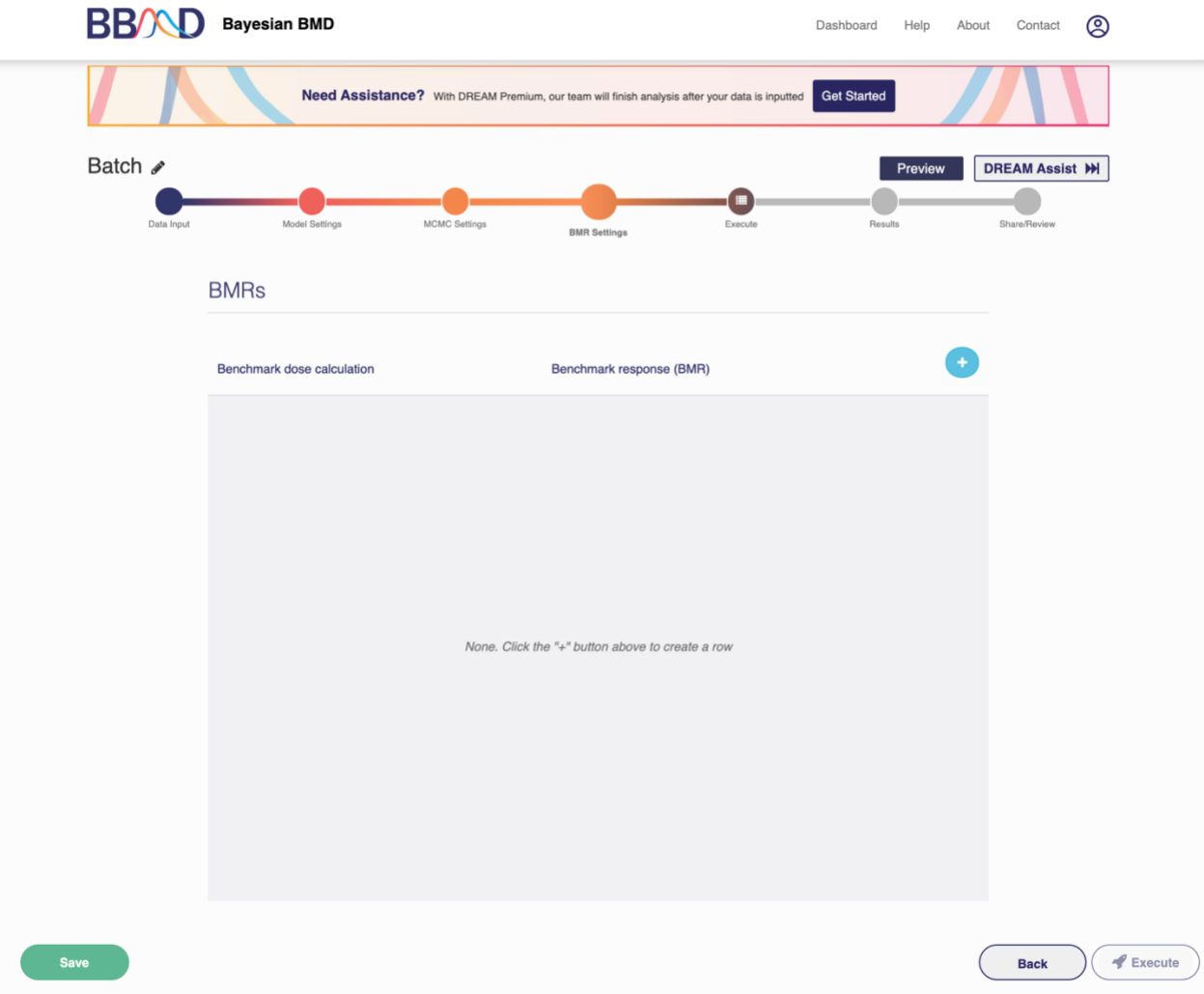


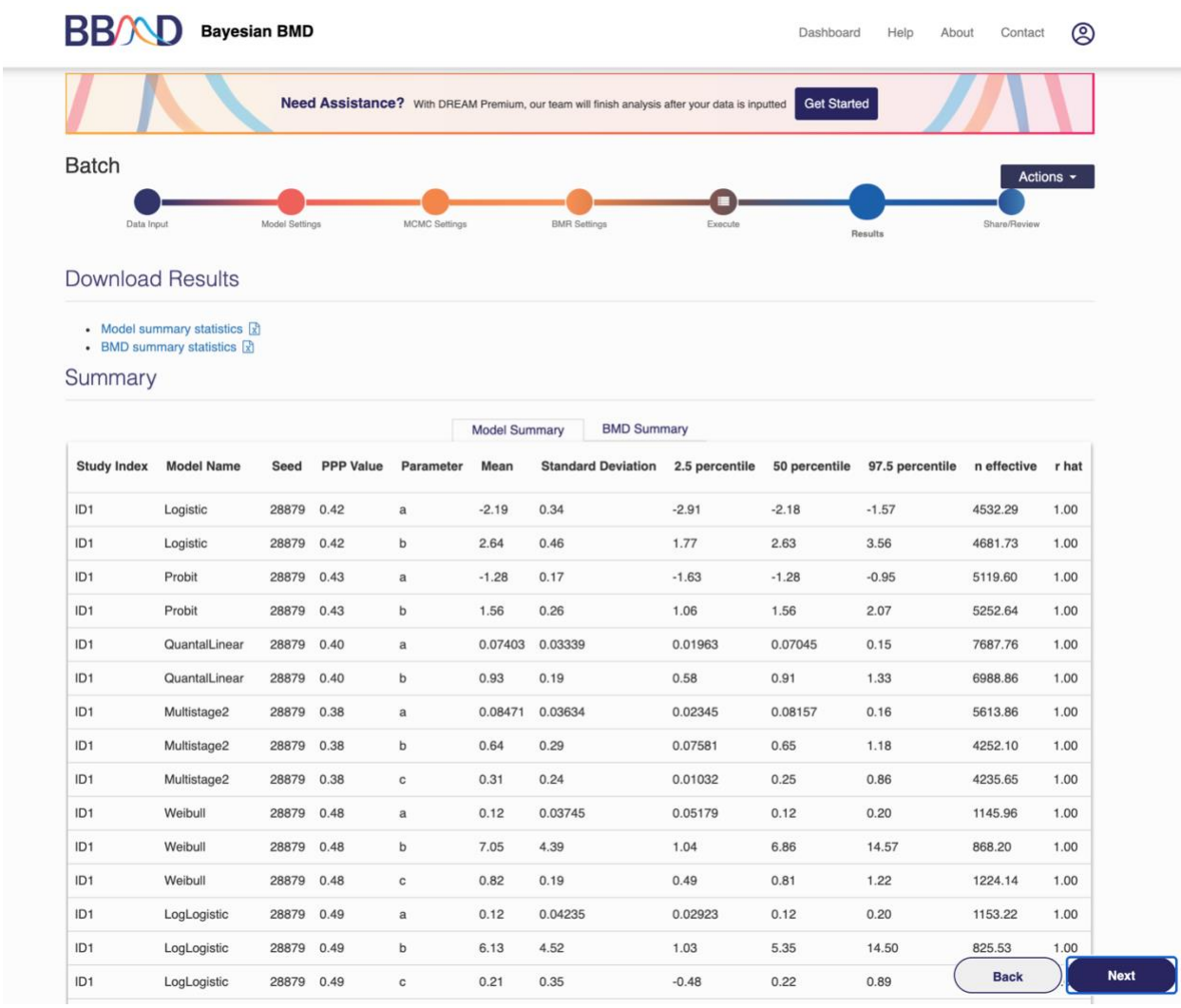
Figure 31. "BMR Settings" tab

E. Execute

While the BMD Batch analysis is executing, you may leave the system. When the analysis is completed, you will receive an email notification, and a link to your analysis. Then you will be able to view, share, and export the results.

F. Results

This tab, shown in Figure 32, displays the model summaries and BMD summaries for the batch analysis. You can either download the Excel spreadsheet versions of the results or view them in the tables on this page.



Study Index	Model Name	Seed	PPP Value	Parameter	Mean	Standard Deviation	2.5 percentile	50 percentile	97.5 percentile	n effective	r hat
ID1	Logistic	28879	0.42	a	-2.19	0.34	-2.91	-2.18	-1.57	4532.29	1.00
ID1	Logistic	28879	0.42	b	2.64	0.46	1.77	2.63	3.56	4681.73	1.00
ID1	Probit	28879	0.43	a	-1.28	0.17	-1.63	-1.28	-0.95	5119.60	1.00
ID1	Probit	28879	0.43	b	1.56	0.26	1.06	1.56	2.07	5252.64	1.00
ID1	QuantalLinear	28879	0.40	a	0.07403	0.03339	0.01963	0.07045	0.15	7687.76	1.00
ID1	QuantalLinear	28879	0.40	b	0.93	0.19	0.58	0.91	1.33	6988.86	1.00
ID1	Multistage2	28879	0.38	a	0.08471	0.03634	0.02345	0.08157	0.16	5613.86	1.00
ID1	Multistage2	28879	0.38	b	0.64	0.29	0.07581	0.65	1.18	4252.10	1.00
ID1	Multistage2	28879	0.38	c	0.31	0.24	0.01032	0.25	0.86	4235.65	1.00
ID1	Weibull	28879	0.48	a	0.12	0.03745	0.05179	0.12	0.20	1145.96	1.00
ID1	Weibull	28879	0.48	b	7.05	4.39	1.04	6.86	14.57	868.20	1.00
ID1	Weibull	28879	0.48	c	0.82	0.19	0.49	0.81	1.22	1224.14	1.00
ID1	LogLogistic	28879	0.49	a	0.12	0.04235	0.02923	0.12	0.20	1153.22	1.00
ID1	LogLogistic	28879	0.49	b	6.13	4.52	1.03	5.35	14.50	825.53	1.00
ID1	LogLogistic	28879	0.49	c	0.21	0.35	-0.48	0.22	0.89		

Figure 32. "Results" tab for BMD Batch analysis

In the Model Summary table, each model has a row for each parameter in the model. The columns of the table are Study Index (dataset name), Model Name, Seed, PPP Value, Parameter, Mean, Standard Deviation, 2.5 percentile, 50 percentile, 97.5 percentile, n effective, and r hat.

The BMD summary table displays the results for the BMDs specified a couple tabs earlier. The table columns are Study Index (dataset name), BMR Type, BMR Value, Model, model prior weight, model posterior weight, BMDL, BMD, and BMDU.

G. Share/Review

The final tab for this analysis is the "Share" tab, shown in Figure 33. By default, all analyses in a personal account can only be accessed by the owner of the account. If you would like to share your analysis with others, you can change the settings from "Private" to "Public", "Organization", or "Dream Assist". The public setting allows you to send the created URL to others for them to access and review (but not edit) the analysis. The 'Organization' setting

allows you to share the link with others for them to edit the analysis. Lastly, the 'DREAM Assist' setting will share this analysis with the DREAM Tech team, who can assist with the analysis. The 'Organization' and 'DREAM Assist' settings are coming soon.

Like in the previous tab, you can also download the summary Excel spreadsheets for this analysis. There are no differences between the spreadsheets available to download from the "Results" or "Share" tabs.

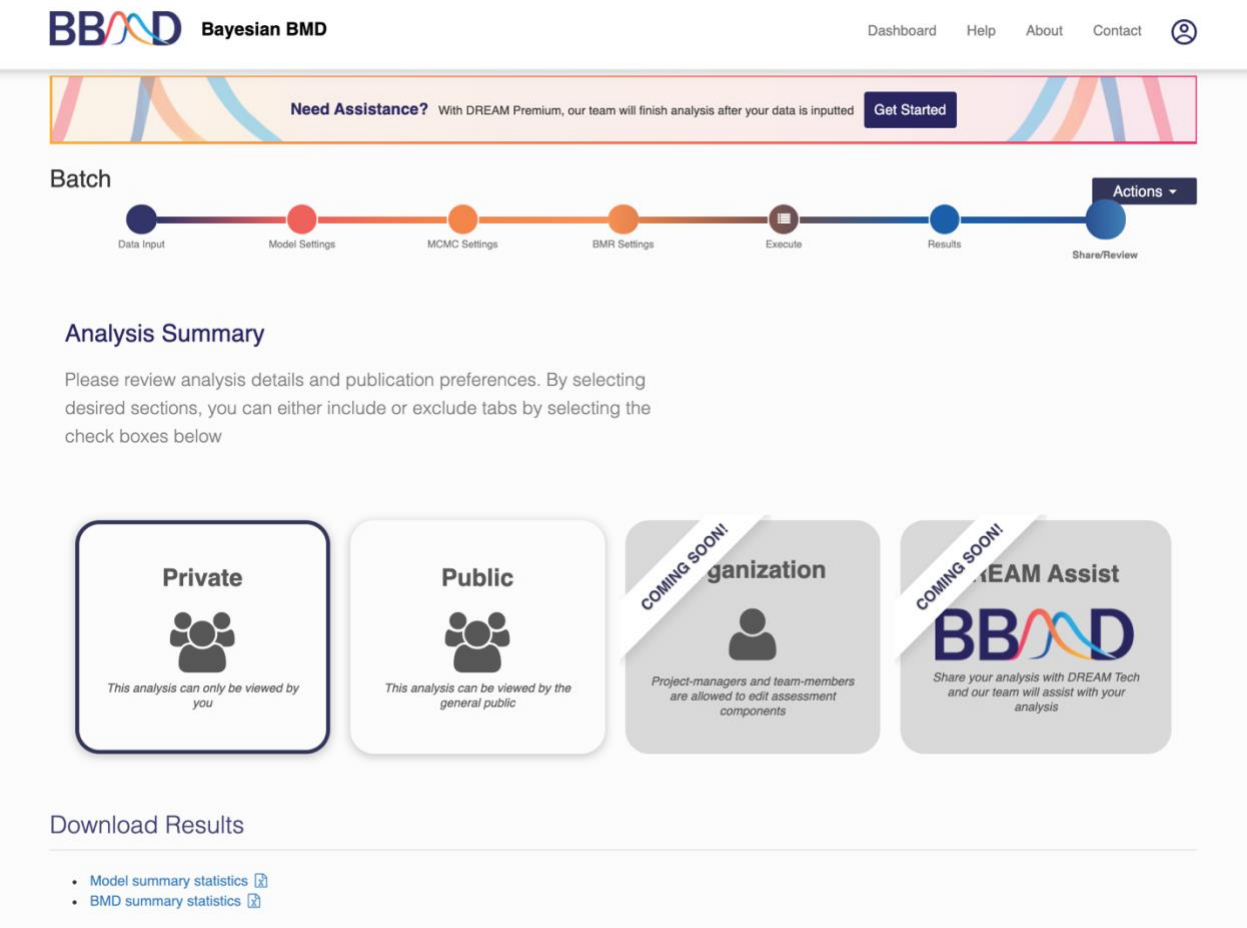


Figure 33. "Share" tab for a BMD Batch analysis

IV. BMD Analysis for Genomic Data

A. Introduction to BMD analysis for Genomic Data

This analysis is designed to perform BMD modeling and estimation for genomic data.

When beginning a new analysis, an automatically generated name "New genomic BMD analysis, *Month Day Year, HH:MM PM*" is assigned to the analysis. You can click the pencil button next to the analysis name, as seen in Figure 34, to make the name more identifiable.

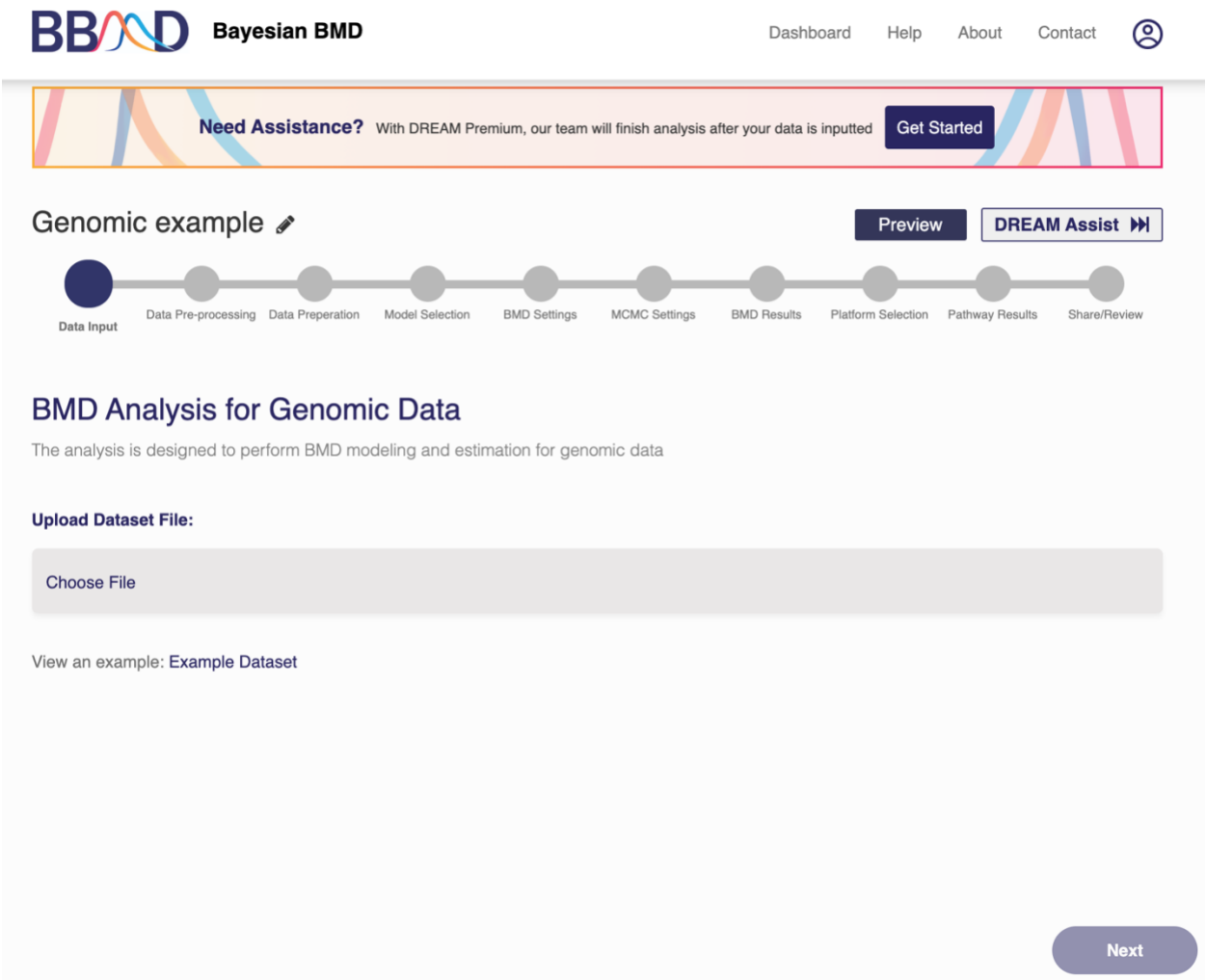


Figure 34. First page of a new genomic BMD analysis

B. Data input

i. How to input data into the system

The first step in performing a BMD Analysis for genomic data is to input your genomic dataset. The dataset needs to be inputted as a '.csv' file and contain the proper columns/rows. The first row should be the Sample IDs, followed by the second row which should contain the dose levels. After that, each row should be labeled with the Gene name. Each column should be a different sample ID and dose level. The table below shows the first few columns and rows of an example dataset.

Example Genomic Dataset

…

SampleID	2D_RG_PLAT_1_093016_G23	2D_RG_PLAT_1_093016_M23	2D_RG_PLAT_1_093016_L21	2D_RG_PLAT_1_093016_M21
Dose	0	0	4.742541	4.742541
ACAA1_48	7.907	8.564	7.721	7.873
CYP2C8_15146	10.334	10.81	10.192	10.192
EREG_21427	4.083	4.266	4.898	4.054
IL1B_3325	3.82	3.322	3.783	3.469

⋮

The row labels "SampleID" and "Dose" must be present in the dataset, otherwise the dataset will be considered invalid.

ii. How to interpret the data input results

After a dataset is successfully uploaded, a summary table, principal component plots, and a density plot will be available to view, as shown in Figure 35. The summary table will show the dataset file name, the number of doses, the dose levels, the number of samples, number of valid genes, and invalid genes (if any were present). An invalid gene is a gene with insufficient data, such as missing response values. If your dataset contained invalid genes, those genes were removed, and you can continue the analysis with the remaining genes.

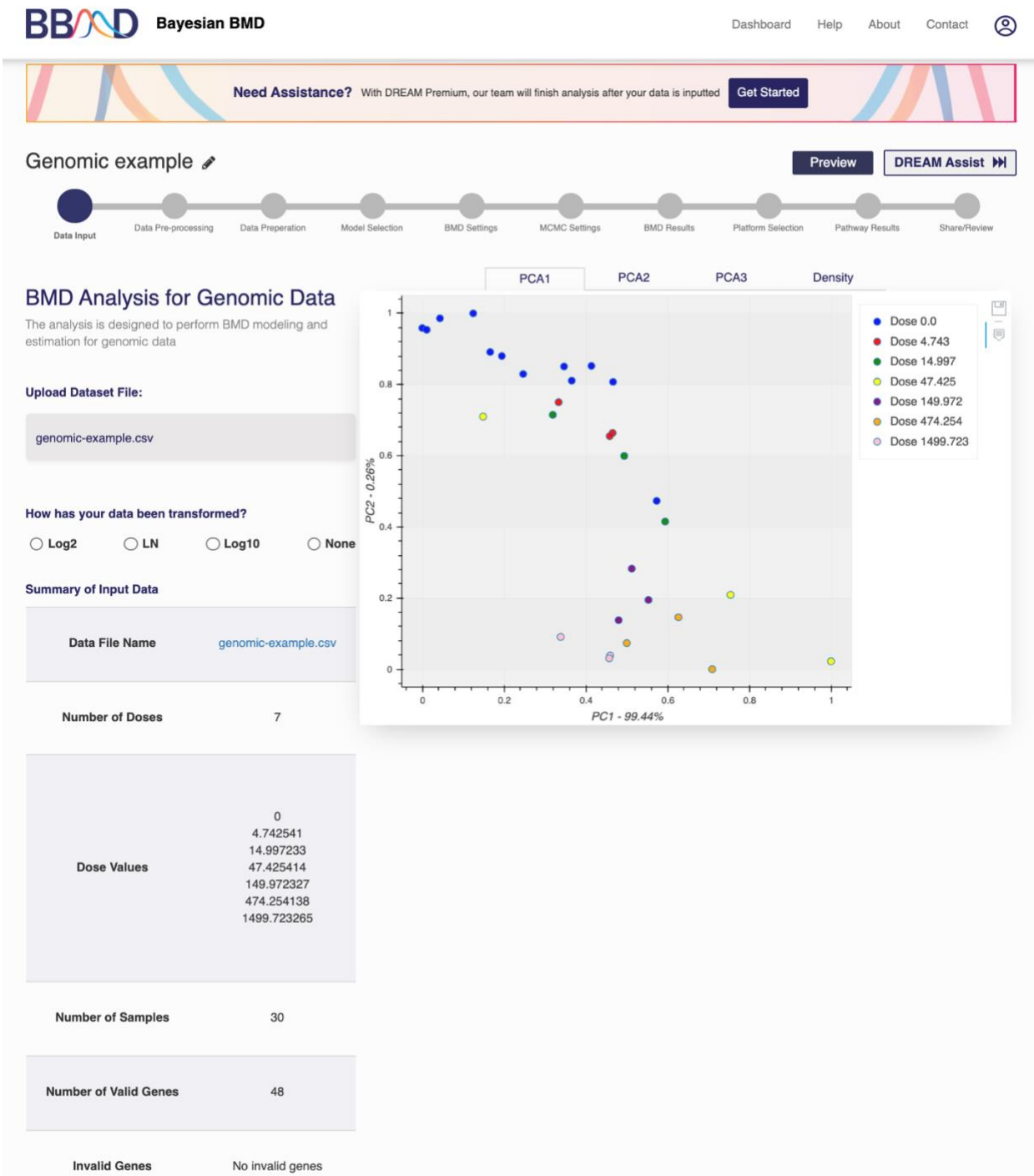


Figure 35. Summary results and the plots available after successfully uploading a genomic dataset

There are also four different plots available to display the relationships found within the data. The first three plots are Principal Component plots, showing PC1 vs. PC2, PC1 vs. PC3, and PC2 vs. PC3. For these plots, each dose level is given a unique color, and you can hover over the points on the plot to show their dose level and sample ID. The PCA plots are intended to demonstrate how well the genomic data clustered before performing an analysis. PCA plots can

be used to identify the outlier samples that seem 'far away' from others in their replicate group. The outliers may reflect some biological interactions that should not be ignored depending on your experimental system. Combining the PCA plots and the density plot to decide if data should be included in the analysis or not. The last plot is the Density plot, which gives you an idea about the signal distribution across microarray chips. Arrays that have very different distributions in the density plot should be checked carefully. Generally, these arrays will show up as problematic and if so, should be removed before analysis.

Before you can continue onto data pre-processing, you must input the type of transform the data has already undergone. The options are $\log_2$, ln, or $\log_{10}$. If the data has not yet been transformed, select "None", and a $\log_2$ transform will be applied. The transform type is necessary for the preprocessing algorithms, so it is critical that the correct option is selected.

Once have finished the data transform step and are satisfied with the summary data, you can click "Next" in the bottom right corner to continue onto the "Data Pre-processing" tab.

C. Data Pre-processing

i. How to perform data pre-processing

The next step in performing a genomic BMD analysis is to use the three preprocessing algorithms to rule out genes that don't fit the specified criteria. This step will reduce the number of genes to be used in the analysis in the BMD analysis and Pathway analysis steps. Figure 36 shows the "Data Pre-screening" tab before the preprocessing algorithms have executed.

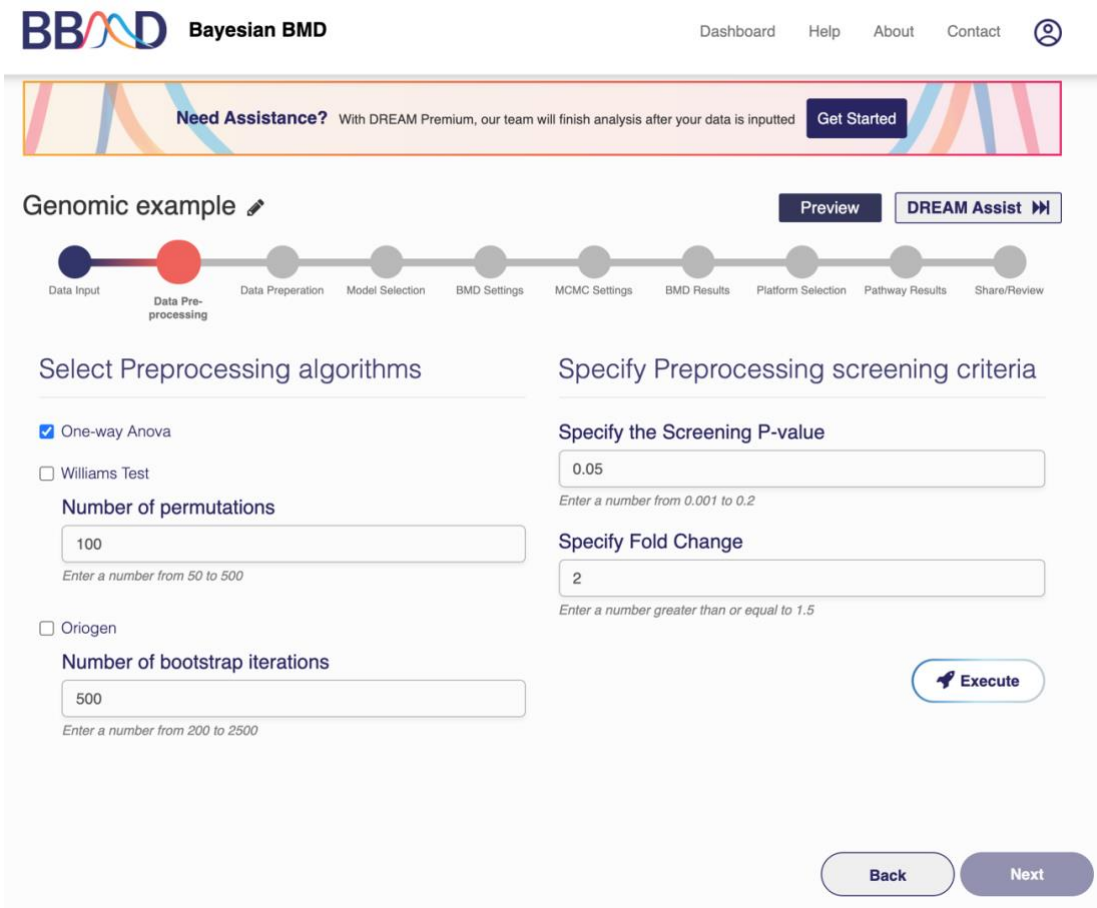


Figure 36. "Data Pre-processing" tab for a Genomic BMD Analysis

Performing data preprocessing takes place in two steps. The first step is to choose the preprocessing algorithms to use and specify the required settings. The second step is to specify the preprocessing screening settings.

The three algorithms used for preprocessing are "One-way Anova", "Williams Test", and "Oriogen". All three of the algorithms use the same screening criteria to select genes that pass, but the screening values are calculated differently in each algorithm, so different genes may pass one algorithm but not another. If you use either "Williams Test" or "Oriogen" you must also specify an additional setting. For the 'Williams Test', you must enter the number of permutations to be used during the algorithm, and for 'Oriogen' you must specify the number of bootstrap iterations to be used during the algorithm. Using a higher number of permutations or bootstrap iterations may allow the algorithms to become more precise but with the cost of taking longer to execute.

The screening criteria that you need to specify is the screening P-value and the fold change. For a gene to pass prescreening, the P-value must be below the value entered in the corresponding text box. The gene's fold change must be above the value entered into that text box.

Once you are satisfied with the selected algorithms, algorithm settings, and prescreening criteria, press “Execute” in the lower right corner. The prescreening algorithms can take a long time to execute, so an email will be sent to your inbox when the prescreening is complete so you can return to the analysis to view the results and move onto BMD and Pathway analyses.

ii. How to interpret the data pre-processing results

When the preprocessing algorithms have finished executing, a summary of the results, and two volcano plots will be displayed, shown in Figure 37. The summary of results will show the number of genes that passed each prescreening algorithm. The volcano plots will show the max fold change vs. the adjusted p-value and unadjusted p-value for each gene that passed the prescreening. These genes will be marked by different colors on the volcano plot, depending on the prescreening algorithm used. You can also hover over the markings on the volcano plot to see the specific gene name so that you can remove or add these genes from your BMD datasets on the next tab.

Note: If the unadjusted or adjusted p-value of a gene is too close to 0, the system my not include these on the plot because Log(p-value) will be undefined.

If you are not satisfied with the prescreening results, you can change the settings and re-execute the prescreening algorithms. Like before, an email will be sent to your inbox when the prescreening has completed.

Figure 37. Preprocessing Summary and Volcano plots

iii. Explaining the algorithms used

Criteria including fold changes, P-values, adjusted P-values are applied to filter gene expression data.

a. <u>Fold change</u>

Fold change (f) is calculated using the following equation. The default value of f is 2, which means dose-response data with f smaller than 2 are filtered.

$$f = \begin{cases} abs \max\left(-\frac{1}{base^{(x_i - x_0)}}\right); \ if \ base^{(x_i - x_0)} < 1 \\ abs \max\left(base^{(x_i - x_0)}\right); \ if \ base^{(x_i - x_0)} \geq 1 \end{cases},$$

where x_i is the response at i-th dose level, x_0 is the response at control level, and base is the log base of data log transformation.

b. <u>P-value and adjusted P-value</u>

P-values are calculated by one way ANOVA and trend test (Williams test and Oriogen). With the P-values, adjusted P-values are calculated by Benjamini-Hochberg methods.

<u>One-way Anova</u>

One way ANOVA is a well-known test to determine whether there are any statistical differences between the means of the experiment group and the control group. The ANOVA produces an F-statistic, the ratio of the variance calculated among the means to the variance within the samples. A higher F-ratio implies that the samples were drawn from populations with different mean values. The F-ratio gives a P-value to determine whether significant differences exist in the experiment and control groups.

<u>Williams Test and Oriogen</u>

We adapt William's trend test (Williams 1971, 1972) and Oriogen (Peddada et al. 2005) to identify genes having a monotonical trend with respect to doses. That is, the maximum likelihood estimate (MLE) of mean response at i-th level is estimated by equation (1) and the test statistic (T) is calculated by equation (2). Permutation and bootstrap methods are applied to calculate the probability (P-value) that T_i (i-th permutation or bootstrap) is larger than T.

$$\widehat{\mu_i} = \begin{cases} max_{1 \leq u \leq i} min_{1 \leq v \leq K} \frac{\sum_{j=u}^{v} n_j \bar{X}_j}{\sum_{j=u}^{v} n_j}; \ if \ increasing \\ min_{1 \leq u \leq i} max_{1 \leq v \leq K} \frac{\sum_{j=u}^{v} n_j \bar{X}_j}{\sum_{j=u}^{v} n_j}; \ if \ decreasing \end{cases}, \quad (1)$$

where $\widehat{\mu_i}$ is the MLE of μ_i, K is dose level, i is the index of treatments group ($i = 1, \dots, K$), n_j is the number of samples at j-th level, and $\bar{X}_j$ is the mean response at j-th level.

$$T = abs \max\left(\frac{\widehat{\mu_i} - \overline{X_0}}{s\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}}\right), \quad (2)$$

where $\overline{X_0}$ is the mean response at the control level, s is an unbiased estimate of within group standard deviation, n_i is the number of samples at i-th level, and n_0 is the number of samples at control level.

D. Data Preparation

After genes have been selected out with preprocessing, you have another opportunity to remove genes from your analysis in the data preparation step. This tab, shown in Figure 38, allows you to add and remove specific genes, and create multiple datasets for the BMD and Pathway analyses.

To create a dataset, follow the three steps below:

(1) Give the dataset an identifiable name in the “Dataset Name” text box.
(2) Choose the preprocessing algorithm which you will select genes from for the dataset.
(3) Add the desired genes to the dataset using the checkbox in the gene’s row. The sliders on the right-hand side of the page can be used to add all genes that fit those three parameters. Even after adjusting the sliders, you can individually add or remove genes using the check box.
(4) When you are satisfied with the dataset, click the green “Create” button on the bottom left side of the page.

You should see the dataset name appear on the left column of the page. If you want to edit this dataset, click on the name, and then change any of the settings previously chosen. After you are finished editing, save your changes by clicking the green “Update” button in the bottom left corner. To add an additional analysis, click the plus icon next to “Datasets” in the left column, then follow steps (1) – (4) again. You may include up to three datasets per genomic analysis. If you need to use more datasets, you can begin a new genomic analysis. If you want to delete a dataset, click the three dots icon next to the dataset name, then click “Delete”. This will remove that dataset from your list to be analyzed. If you had three datasets and then delete one, you can add another.

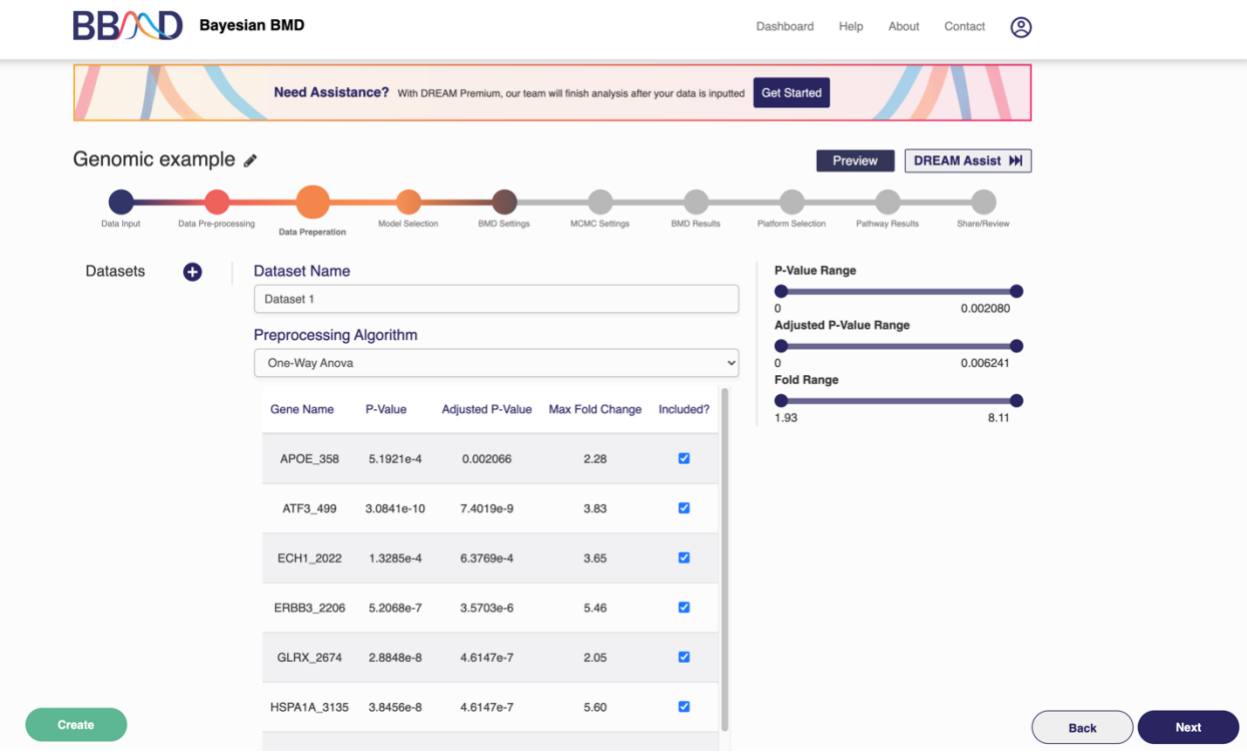


Figure 38. “Dataset Preparation” tab for genomic BMD analysis

E. Model Selection

Before specifying BMD settings, you must select the models to be used in the model fitting part of the analysis.

The available dose-response models are:

1) Linear Model:

$$f(dose) = a + b \times dose, a > 0$$

2) Power Model:

$$f(dose) = a + b \times dose^g$$
$$a > 0, g \geq restriction$$

3) Hill Model:

$$f(dose) = a + \frac{b \times dose^g}{c^g + dose^g}$$
$$a > 0, c > 0, g \geq restriction$$

4) Exponential 2 Model:

$$f(dose) = a \times e^{b \times dose}, a > 0$$

5) Exponential 3 Model:

$$(dose) = a \times e^{b \times dose^g}$$
$$a > 0, g \geq restriction$$

6) Exponential 4 Model:

$$f(dose) = a \times \left(c - (c-1) \times e^{-b \times dose}\right)$$
$$a > 0, b > 0, c > 0$$

7) Exponential 5 Model:

$$f(dose) = a \times \left(c - (c-1) \times e^{-(b \times dose)^g}\right)$$
$$a > 0, b > 0, c > 0, g \geq restriction$$

To include a model in the analysis, click the checkbox on the right side of the model's row in the table displayed in the "Model Selection" tab (shown in Figure 39). When you check the box, the model is automatically added and saved to your analysis. Once you have added all the models you wish to include, press "Next" in the bottom right corner to begin specifying BMD settings.

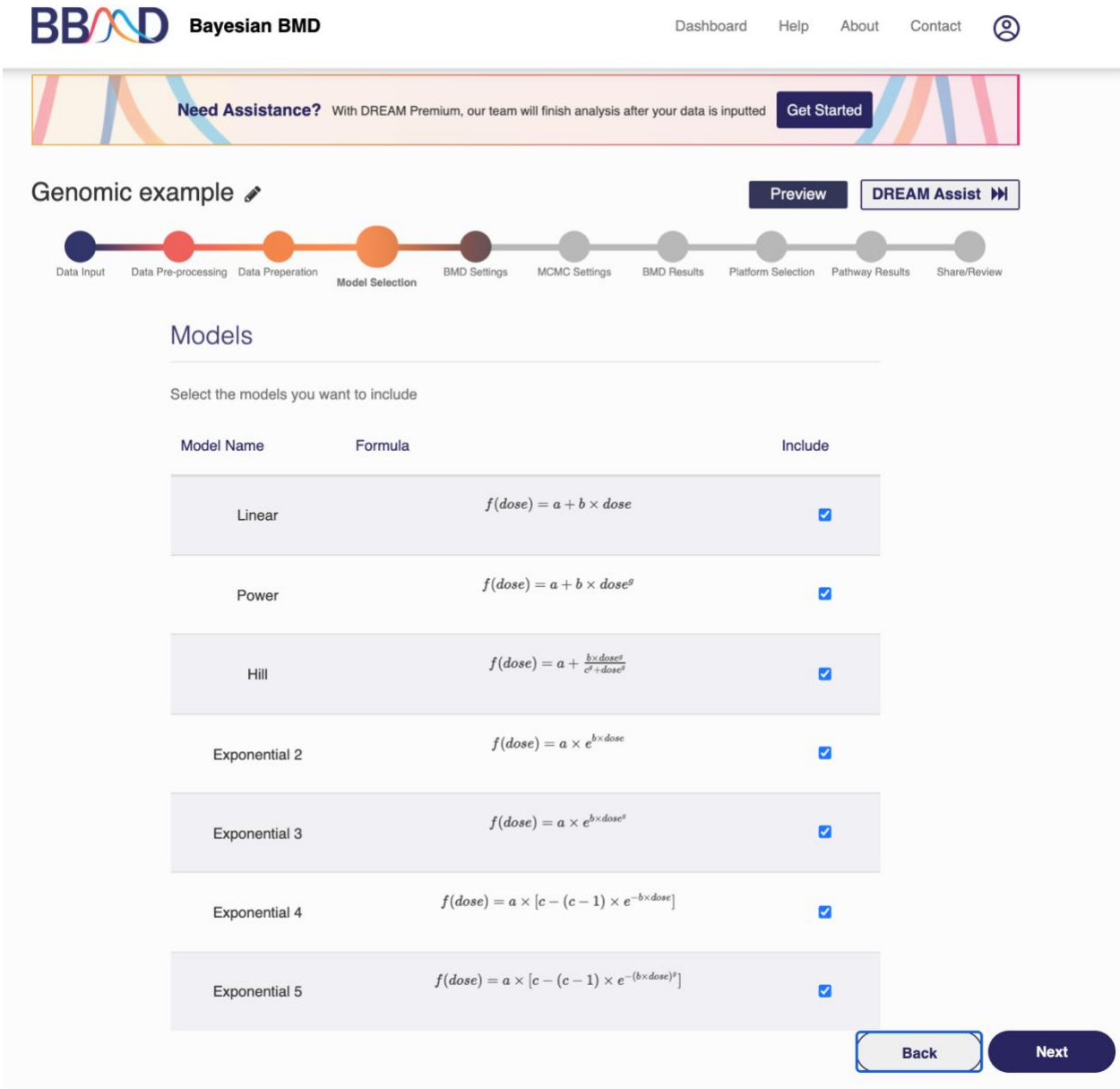


Figure 39. "Model Selection" tab for genomic BMD analysis

F. BMD Settings

i. Genomic BMD analysis steps

The next step in completing a genomic BMD analysis is to specify the BMD settings (the "BMD Settings" tab is shown in Figure 40). Two types of BMRs are available for genomic data: SD Change and Relative Change.

To create a BMD estimate to be analyzed, complete the following four steps:

(1) Choose a name for the estimate. The default name for SD change is "BMD = *BMR Value* SD". The default name for relative change is "BMD = *BMR Value* %". If you would like to use a more identifiable name, enter this into the "Name" text box.
(2) Select a BMR Type from the corresponding drop-down menu.
(3) Specify the BMR value and enter this into the "BMR Value" text box.
(4) Click "Save" at the bottom of the page to save these settings.

When the settings are successfully saved you should see the BMD settings name appear on the left column of the page. To edit these settings, click the name, then follow the previous steps to change any values. Once you have made the desired changes click the green "Update" button on the bottom of the page. If you want to instead cancel these changes, click the "Cancel" button on the bottom of the page. If you would like to add a new BMD setting for this analysis, click the plus icon next to "BMD" on the left side of the page. Three BMD settings are allowed for each genomic analysis. If you would like to delete this settings definition, click the setting's name on the left side, and then click the red "Delete" button on the bottom of the page.

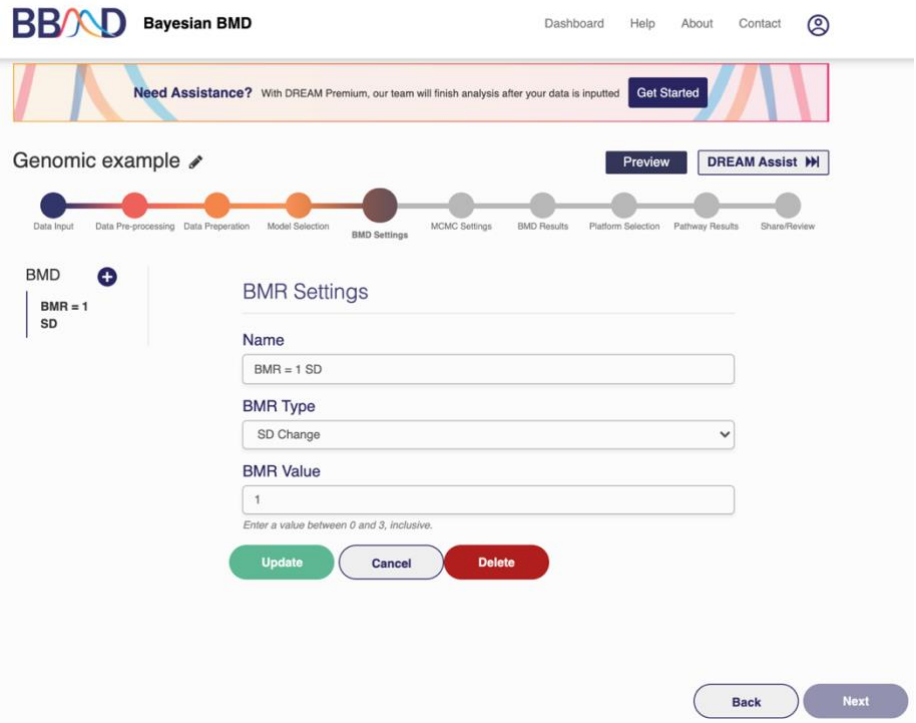


Figure 40. "BMD Settings" tab for genomic BMD analysis

ii. Explanation of the analysis calculation

Two options for defining the BMR values are provided based on the central tendency: a) relative change and b) standard deviation, which are descried below. In BBMD, several BMRs can be defined.

$$f(BMD) \pm f(0) = relative\ change\ \times\ f(0),$$
$$f(BMD) \pm f(0) = \ k \times \text{standard deviation},$$

where $f(0)$ is the estimated response at zero dose, $f(BMD)$ is the response at BMD, relative change (e.g., 10%) and k (e.g., 1) are values defined by user, and standard deviation is estimated by models given observations. As a note, for every model, every posterior sample has an estimate for $f(0)$, standard deviation and therefore a BMD estimated value.

G. MCMC Settings

The final step before executing the BMD analysis is to specify the MCMC settings, the "MCMC Settings" tab is shown in Figure 41.

There are four different values that need to be specified in this tab. The first value is the number of Markov chain iterations, which can be 10,000 to 50,000 (inclusive) iterations per chain. Enter your value into the "Markov Chain Iterations" text box. Next, you need to specify the warmup percentage for each Markov Chain. This is the percentage of iterations discarded from the beginning of each chain; Therefore, those iterations will not be used for estimating model distributions. Put this percentage in the Warmup Percent (%) text box. The third value that needs to be specified is the number of Markov chains used in the analysis. Enter a number 1 to 3 (inclusive) into the "Number of Markov Chains" text box. Each chain will use the number of iterations previously specified. The final value is the random seed which is used for reproducing analysis results. The random seed can be 0 to 99,999 (inclusive). Enter this value in the "Random Seed text box".

Once these values are specified, click "Save" to save the MCMC settings. If you are ready to execute the analysis, press "Execute" on the right side of the page. This will begin the BMD analysis execution. Once the execution has completed you will receive an email notification that your analysis is ready to be accessed. You will also receive a link which will send you back to the analysis to view the BMD results.

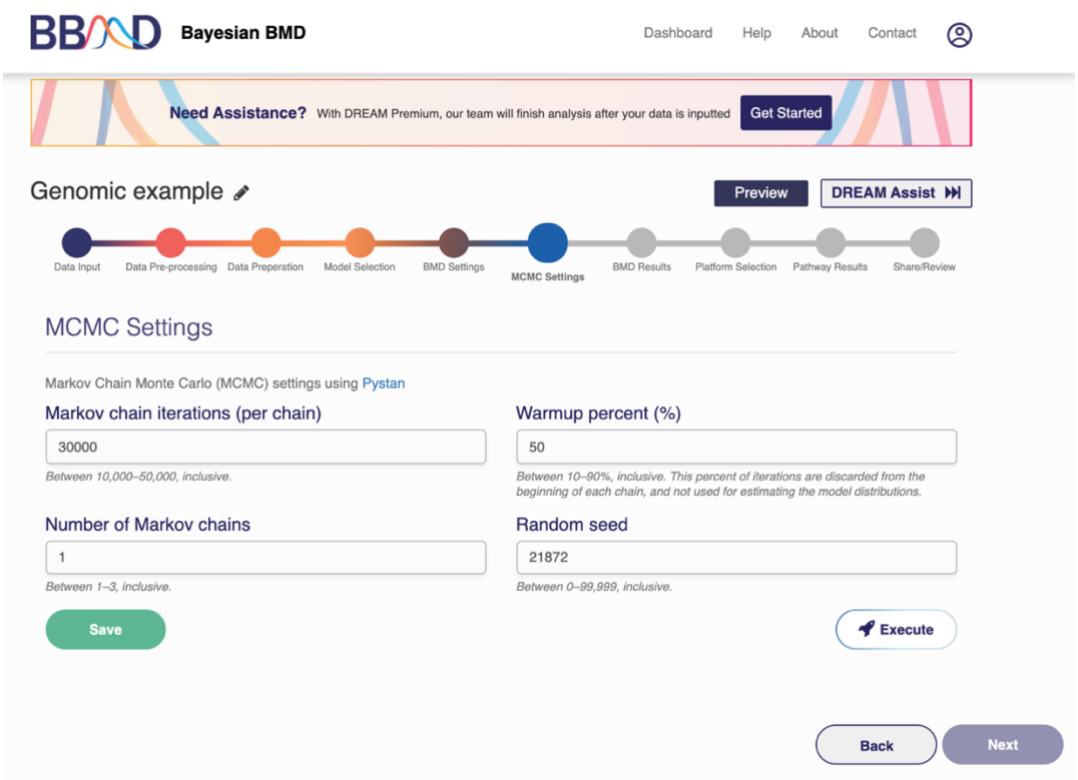


Figure 41. "MCMC Settings" tab for genomic BMD analysis

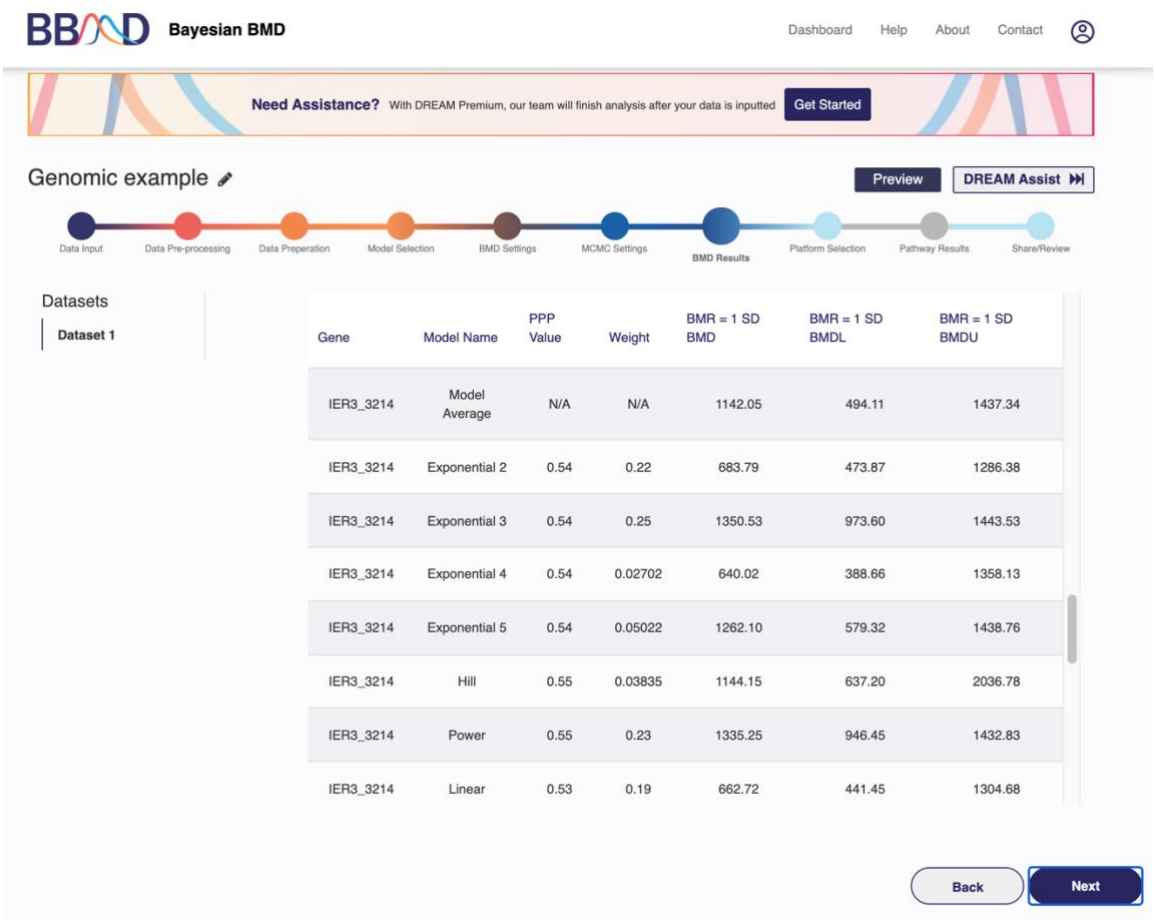
H. BMD Results

Once the BMD analysis has completed, the results will be displayed in the table on the "BMD Results" tab, shown in Figure 42. The names of the different datasets used will be shown on the left side of the page. Click the name of the dataset to view those specific results.

For each gene, the table displays the Model name, PPP value, model weight, and then the BMD, BMDU, and BMDL for each specified BMR. The BMR name will be displayed above the BMD value, and will be the name given when the BMR settings were created.

When you are finished reviewing the results, click "Next" in the bottom right corner to advance to the platform selection tab.

You can also skip the Pathway analysis and go to the "Share/Review" tab instead by clicking the name in the tab bar. This will allow you to export BMD results to an Excel spreadsheet immediately.



Gene	Model Name	PPP Value	Weight	BMR = 1 SD BMD	BMR = 1 SD BMDL	BMR = 1 SD BMDU
IER3_3214	Model Average	N/A	N/A	1142.05	494.11	1437.34
IER3_3214	Exponential 2	0.54	0.22	683.79	473.87	1286.38
IER3_3214	Exponential 3	0.54	0.25	1350.53	973.60	1443.53
IER3_3214	Exponential 4	0.54	0.02702	640.02	388.66	1358.13
IER3_3214	Exponential 5	0.54	0.05022	1262.10	579.32	1438.76
IER3_3214	Hill	0.55	0.03835	1144.15	637.20	2036.78
IER3_3214	Power	0.55	0.23	1335.25	946.45	1432.83
IER3_3214	Linear	0.53	0.19	662.72	441.45	1304.68

Figure 42. BMD Results after executing genomic analysis

I. Platform Selection

In this step, you must select the platform used for the pathway analysis from the drop-down menu. The pathway platforms that are available are shown in Figure 43. When you have chosen a platform, press “Execute” below the drop-down menu. Once the execution has completed, you will be able to advance to the next tab. Click “Next” in the bottom right corner to view the pathway results.

Figure 43. Pathway platforms available to choose from

J. Pathway Analysis Results

i. How to view the pathway analysis results

The "Pathway Results" tab, shown in Figure 44, displays the results for the four different pathway types after completing a pathway analysis.

On the left side of the page, you can select the data set which you would like to view pathway results for. To switch datasets, click on the dataset name. The first section of this page is the pathway type selection. Click the pathway name to switch pathway types. The second section displays the platform summary. Summary information includes: the total number of pathways or total number of genes, number of unique genes, and then the Pathway BMD percentiles. The BMD definitions used for this analysis are the same BMDs that were used for the genomic BMD analysis.

Below the summary section the detailed Pathway BMD results are displayed in a large table. The columns for each table are slightly different, but each table always has a "Gene ID" column or

"Pathway" column. If the text appears blue and underlined in the first column, there is a link to another resource to further analyze that pathway. Similarly, the last columns of the table are always the BMD results columns for each pathway or gene ID. If the text in the results cell is blue and underlined, clicking on it will bring up a plot showing this gene's dose level with whiskers showing the BMDL and BMDU values. If any other text is blue and underlined in the table, hover over those cells because this row corresponds to multiple Probe IDs and Gene IDs.

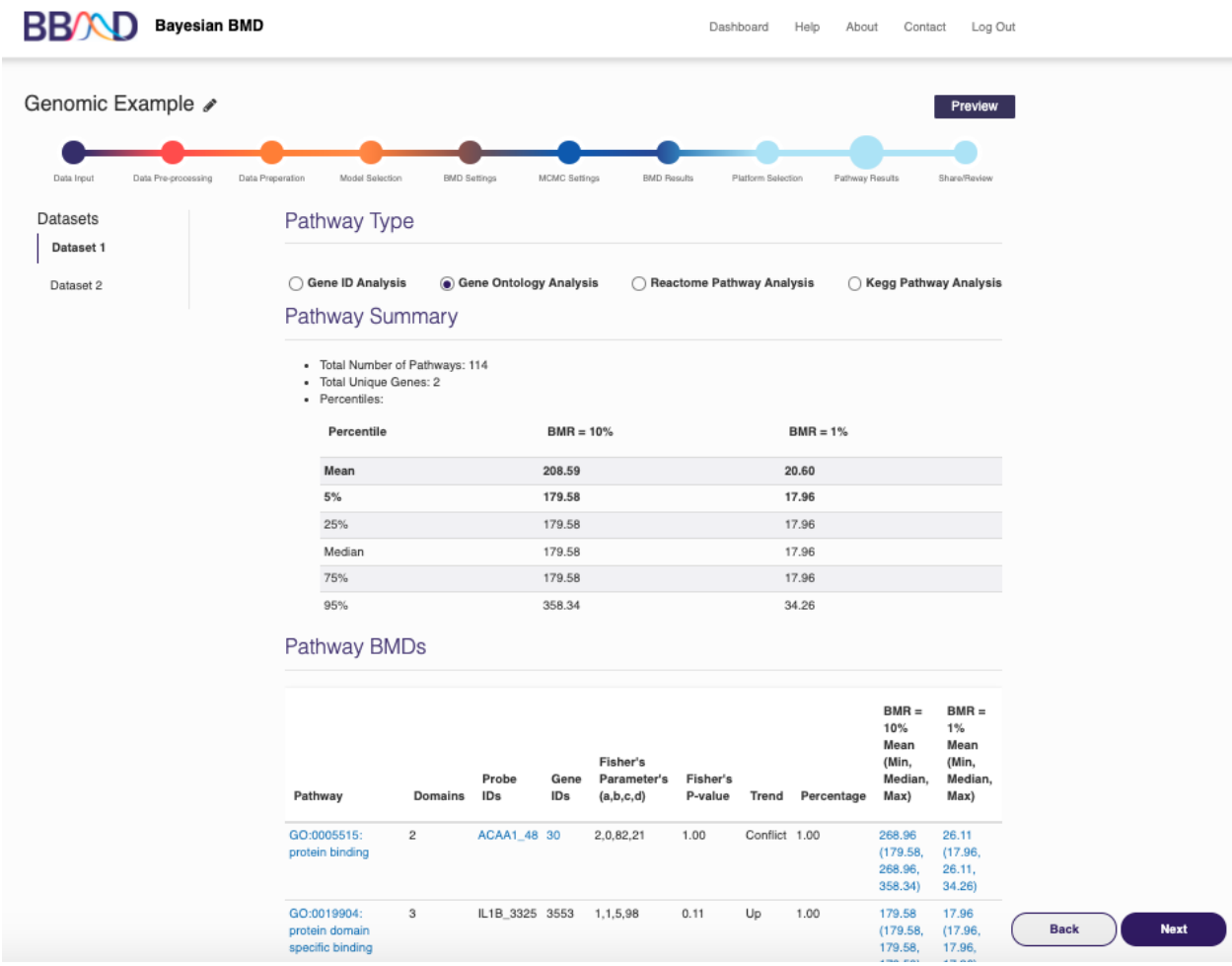


Figure 44. Pathway Analysis Results for "Gene Ontology Analysis"

ii. Algorithms used for pathway analysis

In the pathway analyses, platforms from Gene Expression Omnibus (GEO) are provided and users need to select one that associated with the uploaded genomic data. Four kinds of pathway analyses are provided to classify the BMA BMD analyses into significant pathways based on their NCBI Entrez Gene identifiers: a) Gene ID Analysis; b) GO Analysis (Mi et al. 2019); c) REACTOME (Fabregat et al. 2018) Pathway Analysis; and d) KEGG (Kanehisa and Goto 2000)pathway analysis. Each pathway analysis, the genes were matched to their associated categories, and the minimum, maximum, average, and median BMD were calculated for each category. For each analysis, a pathway summary table and a detailed pathway BMD table are displayed.

As for the four pathway methods, Gene ID Analysis simply translates the probe set identifiers to NCBI's Entrez Gene identifiers. GO Analysis utilizes 'go-basis.obo' and a python package GOATOOLS (Klopfenstein et al. 2018) to group the Entrez Gene identifiers into three sub-ontologies: biological process, cellular component, and molecular function. The 'reactomepy' python module is used to access the REACTOME database, and API request is used to access the KEGG database.

For all the analyses, probe sets that measured more than one gene were removed from analyses. When different probe sets are associated with the same Entrez Gene identifiers, mean values of BMD are taken to represent the BMD value of the Entrez Gene identifiers. In order to determine whether the pathway is significant, P-values and percentages are calculated for each category. P-values are calculated based on Fisher's exact two-tailed test by comparing the numbers of genes with BMD estimates with the numbers of genes without BMD estimates. For each category, percentage is defined as the ratio of the number of genes with BMD estimates that are on this category to the total number of genes that are related to this category. The trend ('Up', 'Down' or 'Conflict') for each pathway is also provided. "Up" indicates > 60% genes in the category show up-regulation, "Down" indicates >60% show down-regulation and "Conflict" indicates neither "Up" or "Down" criteria were met.

K. Share/Review

The final tab for a genomic BMD analysis is the "Share/Review" page, shown in Figure 45. From this tab you can share this analysis with others and export the calculated results.

If you would like to share your analysis with others, you can change the settings from "Private" to "Public", "Organization", or "Dream Assist". The public setting allows you to send the created URL to others for them to access and review (but not edit) the analysis. The "Organization" setting allows you to share the link with others for them to edit the analysis. Lastly, the "DREAM Assist" setting will share this analysis with the DREAM Tech team, who can assist with the analysis. The editable and DREAM settings are coming soon.

You can also export the results of the analysis into Word or Excel formats. Before exporting results, you can customize the parts of the analysis included on the reports. By clicking "show more" on the right side of the page, the customization options will appear. From here you can select which datasets to include in the exported results. If you wish to change the report settings in the future, you can return to this analysis and export a new report. Exporting the results will send a link to your account's email where you can download the reports.

At any time during the updating or reviewing stage, if you want to change to another existing analysis, you can click the "Dashboard" button on the top right corner to switch to the summary page for the existing analyses and access another analysis.

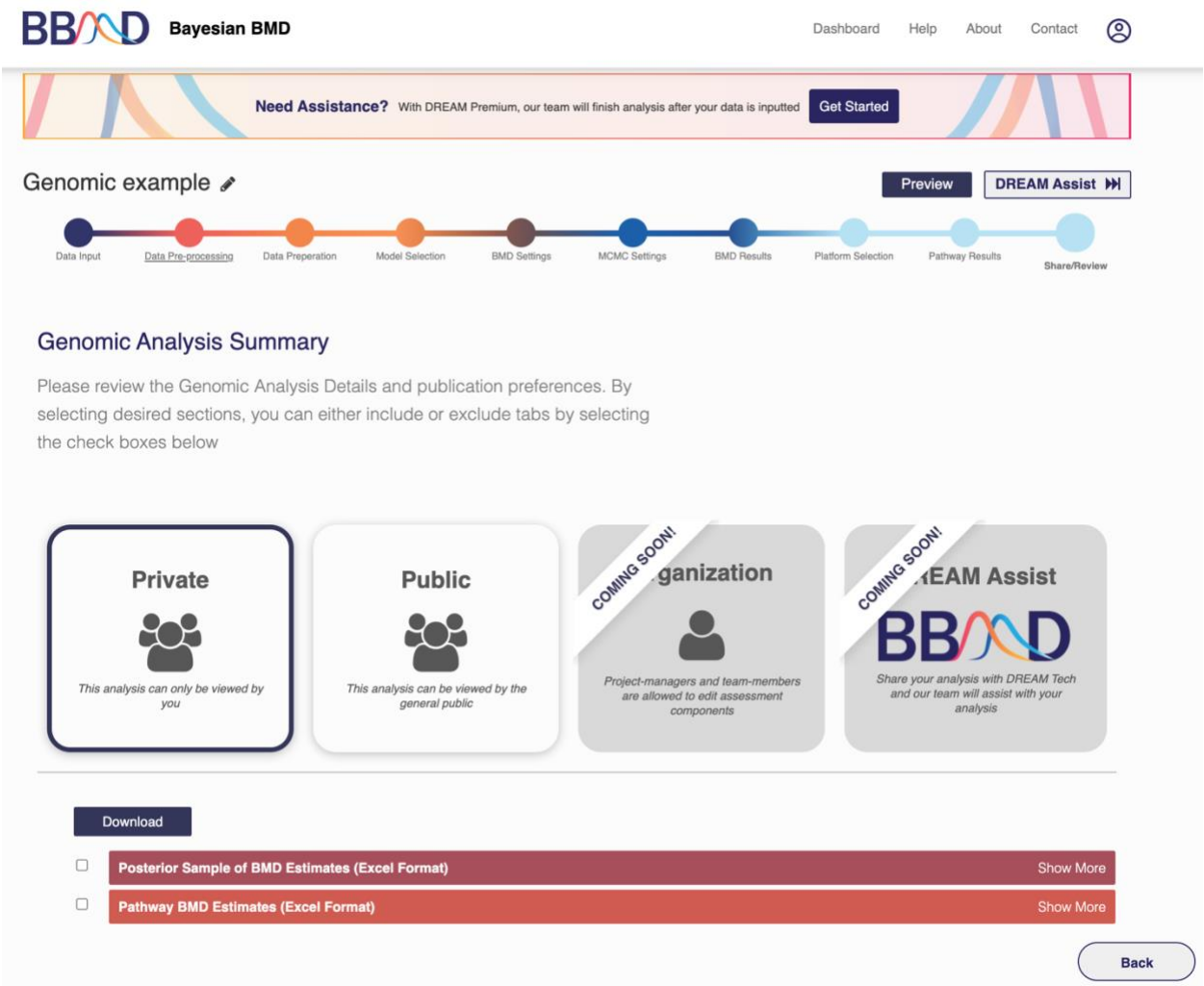


Figure 45. "Share/Review" tab for a genomic BMD analysis

V. Probabilistic RfD Analysis

A. Introduction to the Probabilistic RfD Analysis

This analysis can be used to convert traditionally estimated BMD/BMDL, NOAEL/LOAEL to probabilistic reference dose given some additional input information.

B. Perform a Probabilistic RfD Analysis

An automatically generated name "New pRfD, *Month Day Year, HH:MM PM*", is assigned to the newly started analysis. You can click the pencil button next to the analysis name, as seen in Figure 46, to make the name more identifiable.

This type of analysis has only two different tabs. The first tab is the data input tab where the analysis settings are specified, the other tab is the share tab.

To perform a Probabilistic RfD analysis, follow the 11 steps below on the "Data Input" tab:

(1) Name the RfD analysis in the "RfD analysis name" box
(2) Specify the dose units in the "Dose Units" box (This is not a required step)
(3) (Optional) Choose a random number seed for the "Random number seed" box
(4) Input your prior BMD/BMDL or NOAEL/LOAEL data. If you choose to use a BMD/BMDL analysis as your point of departure selection, you must input the BMD value and BMDL value in the corresponding text boxes. If you choose to use a NOAEL analysis as the point of departure, along with the NOAEL, you must also choose an endpoint type from the drop-down list, and input NOAEL to BMD geometric mean and geometric standard deviation in the corresponding text boxes. Lastly, if you choose to use a LOAEL analysis, in addition to the same settings as a NOAEL analysis, you must also input the LOAEL to NOAEL geometric mean and geometric standard deviation in the corresponding text boxes.
(5) Specify the Allometric Scaling settings. Select a test species from the "Test species" drop-down menu. Specify the test species body weight in the "Test species body weight" box. Choose a human body weight in the "Human body weight" box. Lastly, specify the Allometric scaling exponent mean and standard deviation in the two corresponding boxes.
(6) Give an Animal to Human Uncertainty geometric mean and geometric standard deviation. These two values go in their respective boxes.
(7) Specify Duration of Exposure details to extrapolate non-chronic exposures to chronic exposures. Select the duration of exposure from the drop-down menu. Then give the short-term exposure geometric mean and geometric standard deviation in the corresponding boxes.
(8) (Optional) Add up to two additional uncertainties. Each additional uncertainty requires a name, geometric mean, and geometric standard deviation. These choices go in the respective boxes.
(9) Specify the Human Variability geometric mean and geometric standard deviation in the corresponding boxes.
(10) Give the target population-based incidence in the I* box.
(11) Click the "Save" button to execute the RfD analysis using the settings just specified.

When the analysis is completed, the name will be displayed on the left panel of the screen along with any other analyses. To edit the analysis settings, click the pencil icon on the right side of the page. You can also click the three dots next to the analysis name and choose “Edit RfD”. If you would like to delete this analysis, click the three dots and choose “Delete”. To add a new analysis, press the plus icon above the names of the existing analyses, and then follow steps (1) – (11) again.

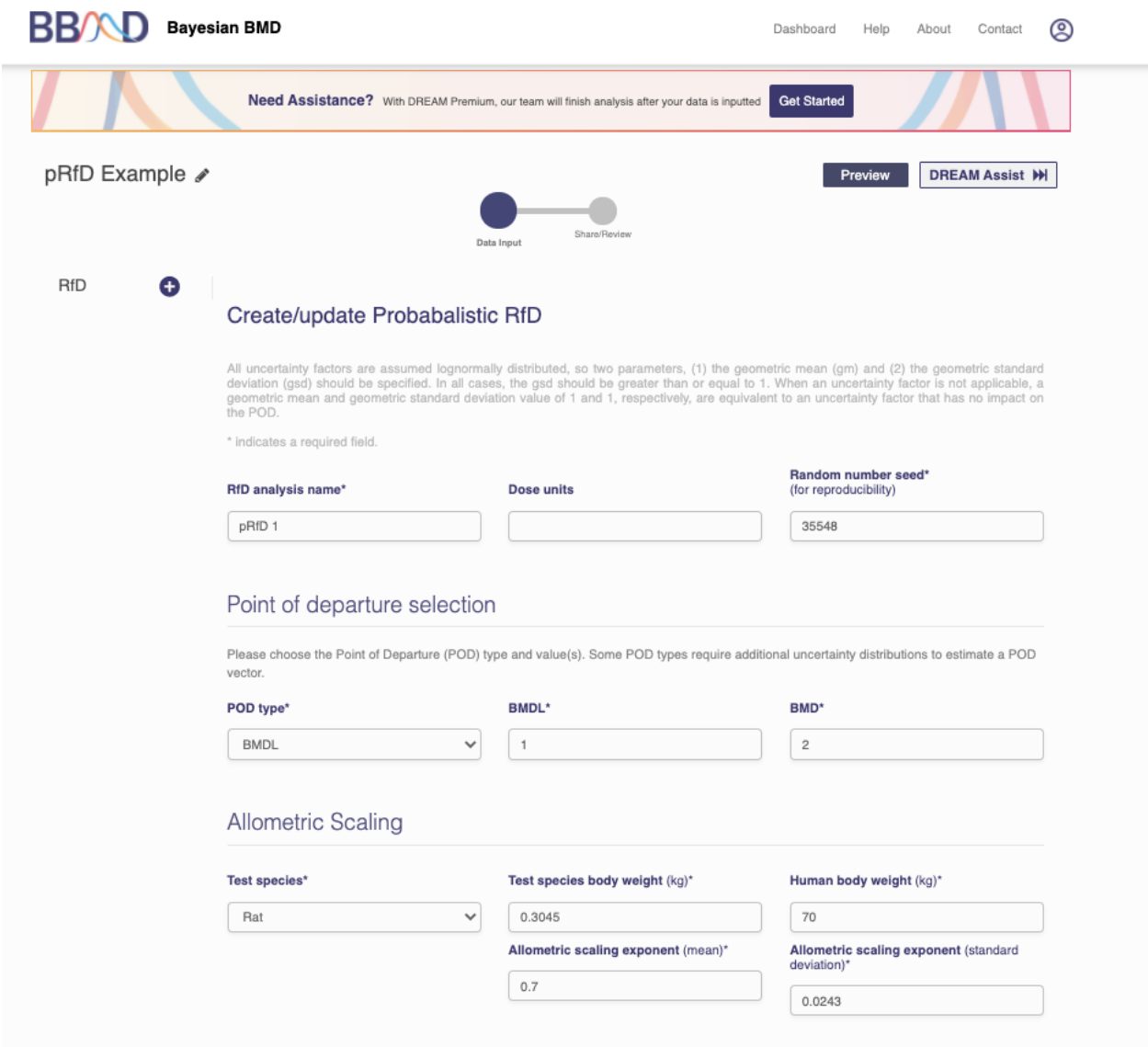


Figure 46. “Data Input” for a Probabilistic RfD analysis

The results, shown in Figure 47, will be displayed also be displayed on the “Data Input” tab. The first section in the results is a summary of the analysis settings. The next section is a table containing the input uncertainty/variability distributions. The next result displayed is the Point of departure (POD) distribution plot. The next plot is the HD_{50}, which is the estimated human

does at which 50% of the population has effects greater than or equal to the target magnitude of effect. The third plot is the HD_M^I, the estimated human dose where the population has 1% incidence of the target magnitude of effect, including inter-individual human variability. The next table contains the data in the plots for percentile and the corresponding POD, HD_{50}, and HD_M^I values. Below this table is the Probabilistic reference dose (RfD) value, defined as the 5th percentile HD_M^I, and the degree of uncertainty (90% CI). The next table is the relative contributions to HD_M^I uncertainty. The final plot shows the target incidence level I* vs. Human dose.

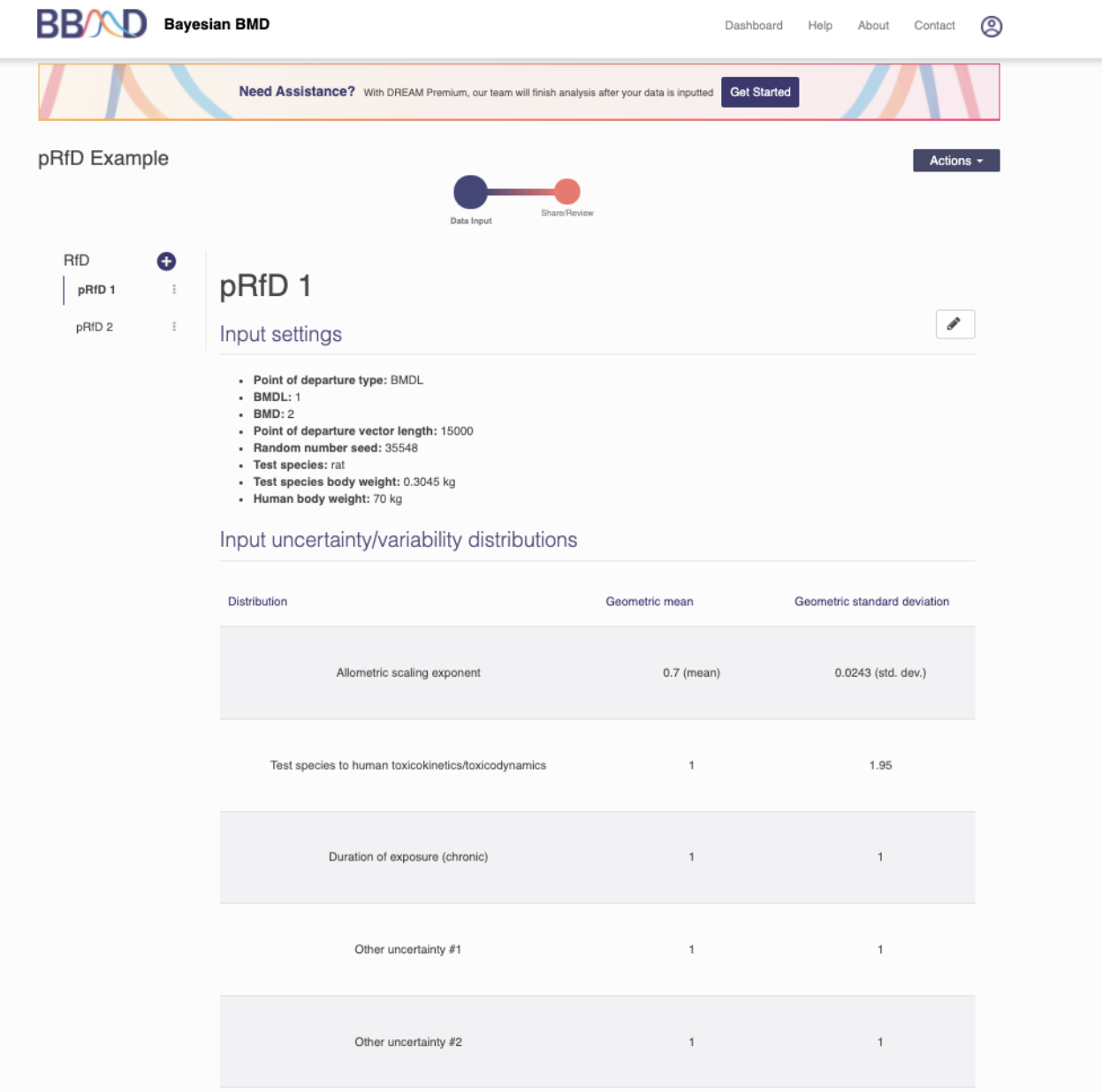


Figure 47. Results from a pRfD analysis

The next tab allows you to share this analysis with others. If you choose “Public”, you can share the analysis link with others for them to view (but not edit). If you choose “Organization”, you can share this analysis with others for them to edit. The last option, “DREAM Assist”, will share it the DREAM Tech team who can help with the analysis.

C. Interpret Probabilistic RfD Analysis Results

With the derived point of departure (POD) from an animal model and the user defiend uncertainty factors, this POD is extrapolated to the equivalent human exposure (e.g., daily oral equivalent dose). Based on the dose response relationship of human extrapolation, the HD50 (50% of the population has effects greater than or equal to the target magnitude of effect) and HD_M^I (estimated human dose where the population has 1% incidence of the target magnitude of effect, including inter-individual human variability) are derived on the RfD results page. The dose-population incidence curve below the summary table gives a visual expression of the statistic values in the HD_M^I table. In addition, the contributions of uncertainty from each individual uncertainty factors are also summarized in the table on the RfD results page.

VI. BMD Analysis for Epidemiological Data (coming soon)

This type of analysis will be coming soon

References

Crump KS. 1995. Calculation of benchmark doses from continuous data. Risk Analysis, 15(1): 79-89.

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.

Slob W. 2002. Dose-response modeling of continuous endpoints. Toxicological Science, 66(2): 298-312.

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

Williams D. 1971. A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics:103-117.

Williams D. 1972. The comparison of several dose levels with a zero dose control. Biometrics:519-531.

Peddada S, Harris S, Zajd J, Harvey E. 2005. Oriogen: Order restricted inference for ordered gene expression data. Bioinformatics 21:3933-3934.

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. Panther version 14: More genomes, a new panther go-slim and improvements in enrichment analysis tools. Nucleic acids research 47:D419-D426.

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. 2018. The reactome pathway knowledgebase. Nucleic acids research 46:D649-D655.

Kanehisa M, Goto S. 2000. Kegg: Kyoto encyclopedia of genes and genomes. Nucleic acids research 28:27-30.