What are the Benchmark Dose Analysis Tests of Interest?

Page Navigation (Quick Links):

Answer: The Benchmark Dose Software (BMDS) was developed by the Environmental Protection Agency (EPA) as a tool to help Agency risk assessors facilitate applying Benchmark Dose (BMD) methods to EPA's hazardous pollutant risk assessments. The software and user guides are regularly updated by the EPA and can be found at the EPA's BMDS website. The information in this document was extracted from the BDMS user guide version 3.0 (EPA document# EPA/600/R-20/216. August 19, 2020).

1. Tests of Interest Table (Tests of Fit)

Tests of Interest			
Test	-2*Log(Likelihood Ratio)	Test df	p-value
1	27.02196286	8	0.00070084
2	1.063745329	4	0.89998092
3	1.063745329	4	0.89998092
4	4.112219646	3	0.24959863

The Tests of Interest table shows the results of four tests based on the log-likelihoods from the Likelihoods of Interest table. The p-values associated with the tests are based on asymptotic properties of the likelihood ratio. Without getting too technical, the likelihood ratio is just the ratio of two likelihood values, many of which are given in the BMDS output. Statistical theory proves that $-2 * \ln(\text{likelihood ratio})$ converges to a Chi-Square random variable as the sample size gets large and the number of dose levels gets large. These values can in turn be used to obtain approximate probabilities to make

inferences about model fit. Chi-Square tables can be found in almost any statistical reference book. Suppose the user wishes to test two models, A and B, for fit. One assumption that is made for these tests is that model A is "nested within" Model B, i.e., that Model B can be simplified (via restriction of some parameters in Model B) in such a way that the simplified model is Model A. This implies that Model A has fewer varying parameters. As an example, consider that the linear model is a "simpler" or "nested" model relative to the power model because the linear model has the power parameter restricted to be equal to 1.

Note: The model with a higher number of parameters is always in the denominator of this ratio.

Suppose that $LL(XX)$ represents the likelihood of model X. Now, using the theory, -2*ln{ L(A)/L(B)} approaches a Chi-Square random variable. This can be simplified by using the fact that the log of a ratio is equal to the difference of the logs, or put,

-2*ln{ L(A)/L(B)} = -2*(ln{L(A)} - ln{L(B)}) = 2*ln{L(B)} - 2*ln{L(A)}.

The values in the Likelihoods of Interest table are in fact the log-likelihoods, as discussed above, ln{$LL(BB)$} and ln{$LL(AA)$}, so this likelihood ratio calculation becomes just a subtraction problem. This value can then in turn be compared to a Chi-Square random variable with a specified number of degrees of freedom. As mentioned in conjunction with the Likelihoods of Interest table, each log-likelihood value has an associated number of parameters. The number of degrees of freedom for the Chi-Square test statistic is merely the difference between the two model parameter counts. In the mini-example above, suppose Model A has 5 fitted parameters, and that Model B has 8. In this case, the Chi-Square value to be compared to would be a Chi-Square with 8 - 5 = 3 degrees of freedom. In the A vs B example, what is exactly being tested? In terms of hypotheses, it would be:

H0: A models the data as well as B

H1: B models the data better than A

Keeping these tests in mind, suppose 2*log{$LL(BB)$} – 2*log{$LL(AA)$} = 4.89 based on 3 degrees of freedom. Also, suppose the rejection criteria is a Chi-Square probability of less than .05. Looking on a Chi-Square table, 4.89 has a p-value somewhere between .10 and .25. In this case, H0 would not be rejected, and it would seem to be appropriate to model the data using Model A. BMDS automatically does the "table look-up" for the user and provides the p-value associated with the calculated log-likelihood ratio having

degrees of freedom as described above. The Tests of Interest table provides four default tests. Associated with each of those tests is a "hover box" that can be accessed to show a summarized interpretation of the test results, which includes EPA's interpretation of the test results (i.e., in relation to p-values that have been selected by EPA). However, the computed p-values are presented so that the users are free to use any rejection criteria they want. Each of the four default tests provided for any of the continuous models is discussed in some detail below.

2. Test 1 (A2 vs R)

Tests the null hypothesis that responses and variances don't differ among dose levels. If this test fails to reject the null hypothesis, there may not be a dose-response.

This test compares Model R (the simpler model) to Model A2. Model R is a simpler A2 (or nested within A2) since R can be obtained from A2 by restricting all the mean parameters to be equal to one another and restricting all the variance parameters to be equal to one another. If this test fails to reject the null hypothesis, then there may not be a dose response, as the inference would be that the simpler model (R) is not much worse than the saturated model. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.05. A p-value less than 0.05 is an indication that there is a difference between response and/or variances among the dose levels and supports a conclusion to model the data. A p-value greater than 0.05 is an indication that the data may not be suitable for dose-response modeling.

3. Test 2 (A1 vs A2)

Tests the null hypothesis that variances are homogeneous. If this test fails to reject the null hypothesis, the simpler constant variance model may be appropriate.

This test compares A1 (the simpler model) to Model A2. Model A1 is a simpler A2 (or nested within A2) since A1 can be obtained from A2 by restricting all the variance parameters to be equal to one another. If this test rejects the null hypothesis, the inference is that the constant variance assumption is incorrect, and a modeled variance is necessary to adequately represent the data. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.05. A p-value less than 0.05 is an indication that the user should consider running a non-homogeneous variance model. A p-value greater than 0.05 is an indication that a constant variance assumption may be suitable for the dose-response modeling.

4. Test 3 (A3 vs A2)

Tests the null hypothesis that the variances are adequately modeled. If this test fails to reject the null hypothesis, it may be inferred that the variances have been modeled appropriately.

Here, the test is one to see if the user-specified variance model, is appropriate. If the user-specified variance model is "constant variance," then Models A1 and A3 are identical; this test is the same as Test 2, with the same interpretation. If the user-specified variance model is nonconstant ($\sigma_i^2 = \alpha * \mu_i^\rho$), this test determines if that equation appears adequate to describe the variance across dose groups. Model A3 is the simpler version of Model A2 obtained by constraining the variances to fit the nonconstant variance equation. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.05. A p-value less than 0.05 is an indication that the user may want to consider a different variance model. A p-value greater than 0.05 supports the use of modeled variance for the dose-response modeling.

5. Test 4 (Fitted vs A3)

Tests the null hypothesis that the model for the mean fits the data. If this test fails to reject the null hypothesis, the user has support for the selected model.

This test compares the Fitted Model to Model A3. The Fitted Model is as simpler Model A3 (or nested within Model A3) because it can be obtained by restricting the means (unrestricted in A3) to be described by the dose-response function under consideration. If this test fails to reject the null hypothesis, the inference is that the fitted model is adequate to describe the dose-related changes in the means (conditional on the form of the variance model; the form of the variance model is the same for the Fitted Model and Model A3). Failure to reject the null hypothesis is associated with the inference that the restriction of the means to the shape of the dose-response function under consideration is adequate. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.1. A p-value less than 0.1 is an indication that the user may want to try a different model (i.e., the fit of the Fitted Model is not good enough). A p-value greater than 0.1 is an indication that the Fitted Model appears to be suitable for dose-response modeling.