MUT/2017/03

COMMITTEE ON MUTAGENICITY
OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT (COM)

Quantitative approaches to the assessment of genotoxicity data II – Evaluation of benchmark dose software.

Evaluation of benchmark dose software paper for discussion with Quantitative approaches to the assessment of genotoxicity data II paper (MUT/2017/02)

Secretariat

February 2017

COMMITTEE ON MUTAGENICITY
OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT (COM)

Quantitative approaches to the assessment of genotoxicity data II

Evaluation of benchmark dose software

1) There are two principle free software packages which are widely available and used for the derivation of benchmark dose (BMD) values for point of departure (POD) determination in toxicology risk assessment. They have been developed for generic use across any kind of toxicity endpoint, in vivo and in vitro, mammalian, human and environmental endpoints as essentially they fit a mathematical model to dose/concentration-response data. Both programmes can be used to evaluate quantal and continuous data and are able to combine analysis for the same endpoint from different studies or groups. It is possible to compare dose responses across different covariates – for example different sex, different study duration, and different tissues (is this true of BMDS?).

2) BenchmarkDose Software (BMDS) was developed by the US Environmental Protection Agency (EPA) in order to standardize approaches to evaluating dose response assessments. The current version is 2.6.0.1 (05/02/17). The software has over 30 different models or model variants which are appropriate for the analysis of quantal data, continuous data, nested developmental toxicology data, multiple tumour analysis, and concentration-time data. The software is freely available on the EPA website and there are training webinars and extensive guides for its use. https://www.epa.gov/bmds

3) The PROAST software package has been developed by the Dutch National Institute for Public Health and Environment (RIVM). The present version is 38.9. A comprehensive discussion of the software is available in the EFSA scientific opinion (2009 – appendix p47-72). PROAST requires the R package to run (available free).

The software is available for download on the RIVM website http://www.rivm.nl/en/Documents_and_publications/Scientific/Models/PROAST

It is noteworthy that both software packages undergo continual updating.

4) Davis et al (2011 = annex), a group from the EPA, describes in detail the BMDS software, its methodology and application and compares it to the PROAST software,. The section evaluating the differences between the EPA and EFSA guidance is reproduced below. They draw attention to the use of the 5% response level by EFSA, and argue that for continuous endpoints, a % value can correspond to different degrees of response depending on the endpoint. A one control standard deviation is recommended by EPA.

> *'The guidance promulgated by EFSA differs from U.S. EPA guidance in a number of ways. The first is EFSA's recommendation for selection of the BMR for continuous endpoints (EFSA's and U.S EPA guidance are similar for dichotomous BMRs). EFSA recommends that a 5% response level is usually satisfactory for continuous data, as it is usually within the observed range of the data and should provide BMD and BMDL estimates that are not overly model dependent, based on the findings of Woutersen et al. (2001) and Sand et al. (2006). The U.S. EPA discourages using a percentage change as the basis for a BMR for continuous endpoints without a biological basis to do so; the same percent change can represent very different degrees of response for different endpoints. U.S. EPA's guidance instructs that a BMR of 1 control standard deviation is a more appropriate BMR for continuous endpoints because it takes into consideration the distribution of the data and is more comparable to the 10% extra risk BMR suggested for dichotomous endpoints.*
>
> *The second way in which EFSA guidance differs from U.S. EPA guidance is in how models are judged regarding fit. EFSA guidance for model fit involves two principles: deciding which model fits best within a "nested" family of increasingly complex models and then a determination of overall goodness-of-fit. Both principles are based on the likelihood ratio test. In the PROAST software there are three families of nested models: the Exponential models (also found in BMDS) and Hill models (only the most complex, four-parameter, form is available in BMDS) for continuous endpoints and the linearized. Multistage models for dichotomous endpoints. For dichotomous endpoints, PROAST also contains all of the dichotomous models available in BMDS. In order to select the best model within a family of models, more complex models must be compared to the corresponding simpler models in order to determine whether the addition of extra parameters significantly improves the model fit. This is done in a step-wise fashion until the most "optimal" (parsimonious) model has been selected. Once the best model from a nested family has been chosen, that "fitted" model and any other singular models included in the analysis is then compared to the "full" model to determine goodness-of-fit. The full model is the model that perfectly fits the means (continuous data) or incidences (dichotomous data) at all dose levels. The*

U.S. EPA BMDS reports p-values derived from likelihood ratio test results (between "fitted" and "full" models) 6 and includes a nested analysis of Exponential models similar to that which is performed in PROAST. However, U.S. EPA recommends that each model fit be judged independently (before model comparisons among models of a nested family). In addition, BMD modeling is largely considered a curve-fitting exercise involving a suite of models, and U.S. EPA (2000) recommends that α=0.1 be used to compute the critical value for goodness-of-fit, instead of the more conventional value of 0.05 used by EFSA.7 Finally, EFSA does not support the use of χ2-scaled residuals to assess local fit, a factor that the U.S. EPA considers important to ensure that the models of interest are providing good local fit, especially in the low-dose region.

Final model selection in PROAST using EFSA's guidance is solely dependent on the lowest BMDL. Unlike U.S. EPA guidance, no consideration is given to relative model fit or the divergence of BMDL results at this point. Pursuant to U.S. EPA's guidance, the "lowest BMDL" criterion is only used when BMDLs are considered to be sufficiently divergent indicating a high degree of model dependence (see BMD analysis step 4). When BMDLs are not sufficiently different and model dependence is unlikely, AIC values the Akaike Information Criterion - should be used in order to determine which model most parsimoniously fits the data.

Additional research and analysis is needed in order to determine how the differences between PROAST and BMDS guidance affect dose– response analyses. As the state-of-the-science evolves, so will the specific guidance promulgated by domestic and international health agencies, and some harmonization of methods can be reasonably anticipated. For example, currently within BMDS, the exponential models can be tested in a nested fashion, and the provided p-values, based on the likelihood ratio tests, can be used to make model selection'.

5) A comprehensive evaluation of the development and application of BMD approaches was undertaken by EFSA (2009). EFSA scientific opinion: use of benchmark dose approach in risk assessment - This report has an appendix, 'software for BMD analysis' in which it describes in detail the EPA benchmark dose software (BMDS) (Version 2.0) and the PROAST software(VERSION NOT GIVEN) .

6) Their summary of differences between BMDS and PROAST is as follows::

A comparison of the differences between the BMDS and PROAST software, taken from is provided in Table A15.

For continuous data, the default assumptions regarding the distribution of the data differ between BMDS and PROAST. As a default, data are assumed to be normally distributed in BMDS while they are assumed to be log-normally distributed in PROAST. If this is the only difference (i.e. the same model, BMR, and other settings), this should result in only slight differences in the BMD and BMDL.

The procedure in PROAST for fitting the family of exponential models and determining the most appropriate member using likelihood ratio tests is available in the BMDS, but at this point only a beta version of this approach has been released.

Appendix – Software for BMD analysis

There are differences with respect to the Hill model family. In PROAST this family is defined as a nested family of models, analogous to the exponential family. In BMDS the Hill model is only available as the 4 parameter model within the Hill family of models.

For continuous data, the variance can be either specified as constant or non-constant in BMDS, while PROAST always uses a constant coefficient of variation. A constant coefficient of variation is a special case of the non-constant variance model in BMDS, i.e. the case when the parameter "rho" equals 1.

In BMDS, several ways of defining the BMR are available for continuous data, whereas, in PROAST, only the options called "Rel. Dev." and "Std. Dev." in BMDS, are available.

For most models in BMDS, only the lower bound of the confidence interval is calculated, i.e. the BMDL, while both the lower and upper bound are computed by PROAST.

For analysis of quantal data, BMDS and PROAST are essentially the same.

Table
Comparison of BMDS and PROAST software.

	BMDS	PROAST
Environment	Can be run immediately as an executable in Windows	Splus or R software (free) is required
Time to get started	Short. Can be used immediately upon download	Steeper "learning curve." Requires basic knowledge of Splus or R
ser Interface	Fully Windows based	Ordered process of answering multiple choice questions; Graphical User Interface available only for continuous data
Models:		
Continuous	Yes	Yes
Dichotomous	Yes	Yes
Nested continuous	No	Yes
Nested dichotomous	Yes	Yes
Categorical	No	Yes
Global goodness of fit test		
Continuous	Likelihood ratio test	Likelihood ratio test
Dichotomous	χ^2 *p*-value	Likelihood ratio test
Model selection criteria		
Model Dependence[a]	Lowest BMDL	Lowest BMDL
No Model Dependence[a]	Lowest AIC	Lowest BMDL
Confidence interval calculated using profile likelihood	Yes	Yes
Confidence interval calculated using bootstrapping	No	Yes
Covariates	No[b]	Yes
Automatic model fitting for nested models	Yes	Yes
Graphic output	Yes	Yes

Reference to the use of ISTD vs BMD 10 or 5%

6) More recently, EFSA have released a new Guidance Document ' Use of benchmark dose approach in risk assessment' for consultation (released 5 December 2015, comments by ???). It is stated that the main modifications to the 2009 guidance is to the section on how to apply the BMD approach.

The summary states

'Model averaging is now recommended as the preferred method for calculating the BMD confidence intervals, while acknowledging that the respective tools are still under development. As these tools may currently not be easily accessible to all, the simpler approach of selecting / rejecting models is still considered as a suboptimal alternative. The set of default models to be used for BMD analysis has been reviewed and a new criterion (the Akaike Information Criterion - AIC) has been introduced instead of the log- likelihood to characterise the relative goodness of fit of different mathematical models to a dose response dataset. A flow chart has also been inserted in this update to guide the reader step-by-step when performing a BMD analysis, as well as a template for reporting a BMD analysis in a complete and transparent manner.'

7) The document also includes a section on parameter constraints in modelling continuous or quantal data

To avoid the models having undesirable properties, certain constraints are imposed on the model parameters. For instance, since continuous responses are usually positive, the background response parameter (a) is constrained to be positive in the continuous models. In quantal models it is constrained to be between 0 and 1 (i.e., 0% and 100% response). Next to the parameter constraints shown in Table 3, an additional parameter constraint has often been applied in practice (US EPA, 2012). This constraint relates to the shape parameter that can be viewed as reflecting the steepness of the curve, i.e. parameter c in the quantal dose-response models (c>1), and parameter d in the continuous (exponential and Hill) models (d>1). The rationale behind this constraint was to avoid that the dose-response would have infinite slope at dose zero. In most models, this may be achieved by constraining the steepness parameter to be larger than one (rather than larger than zero). At first sight, this appears to be a reasonable restriction from a biological point of view. However, as shown in Slob and Setzer (2014), this constraint is based on a false argument and contradicted by real dose-response data. One way to see this is by imagining a study with eight doses between 50 and 0.000005 mg/kg, dose spacing being a factor of 10. The study results in the (quantal) responses are illustrated in Fig. 7. In the upper panel, the responses are plotted against dose. Fitting a model would result in the steepness parameter c being smaller than one, i.e. the dose- response curve has infinite slope at dose zero. In the lower panel, however, the same data are plotted against log-dose, which shows that there is in fact a large range of doses with virtually no change in response. The constraint that the steepness parameter should be larger than one is inappropriate and should not be applied, as it may lead to

artificially high BMDLs. A practical consequence of omitting this constraint is that the BMDL in some cases can be much lower as compared to analysis where the constraint is applied. Section 2.5.7 discusses how to deal with BMDLs that are orders of magnitude lower than the associated BMDUs.

8) A need to establish guidance specifically for the use of BMD approach using human data was also identified It was reiterated that current toxicity test guidelines should be reviewed given the wide application of the approach (ie to review the number of dose groups and numbers of animals within groups). There was no change to the section 2.4.2 risk assessment of substances which are both genotoxic and carcinogenic; that the BMDL10 for carcinogenicity should be useful as the reference point for establishing the MOE

9) A technical report for EFSA scientific panels and units for the use of the BMDS and PROAST software packages was published in 2011 (EFSA 2011). This was based on an EFSA led workshop and is a summary of the exercises, presentations and discussions from that event. Version 2.1.1 of the EPA BMDS software and version 26.0 of the RIVM PROAST software were used. Participants were given a number of toxicology data sets to model – this included quantal and continuous data sets (Quantal data were the presence of urinary bladder stones following administration of melamine to rats in a 13 week feeding study – two separate studies, continuous data were from a rat study of an organophosphate ester and AchE activity); Different models within each software programme were executed (eg Exponential and Hill), the focus of the analyses appears to have been the BMDS software (i.e. not every data set was run through both programmes)

10) The main differences between the two software packages were described (*3.3. Comparison of the software packages)*

- *BMDS uses a window driven environment and is therefore more user-friendly than PROAST that uses the R-software environment. In the present version of PROAST the user needs to answer multiple choice questions, but RIVM is currently working on a graphical user interface for PROAST, i.e. a window driven environment.*
- *PROAST uses the lognormal distribution as the default while BMDS has the normal distribution as the default setting.*
- *BMDS does not at the moment allow for covariates to be included in the analysis while PROAST does.*
- *BMDS is not suitable for studies with a large number of individual data points as there is a limit in the number of rows in the data file; the software is therefore of limited use for human studies.*

• *The same exponential family of models can be fitted in BMDS and in PROAST. The BMDS software gives the outcomes for each model of the family, leaving it to the user to select the "optimal" model, while PROAST selects it automatically.*

11) The overall discussion highlighted; the group agreed with the general consensus that the use of BMD is preferable to NOAEL for deriving POD's. It was noted that a description of the constraints and assumptions that are used for any dose response assessment need to be reported (eg p-value for accepting a model by EFSA is 0.05, whilst the EPA use 0.1). It was clarified that the P-value of the goodness-of-fit test should be used in a relative rather than in an absolute sense, i.e. to distinguish the 'good' from the 'bad' models.

12) It is noted that the software is constantly being updated. The use of the OECD guidelines for testing chemicals was also discussed – that the precision of the analysis can be improved by increasing the number of groups and using less animals per dose group (as discussed previously at COM)>

13) A development of the PROAST software to evaluate genetic toxicology data is available as an online tool, MUTait. This includes the capability to evaluate dose response data (Avancini et al 2016)

There are two major differences in the default approaches:

1. BMDS (EPA) uses the best transformation of the response data for the analyses, whereas PROAST (RIVM) uses the default assumption of a log-normal distribution.
2. BMDS uses the 1 standard deviation (1SD) above the background as the benchmark response /critical effect size (BMR/CES) for continuous data, whereas PROAST uses a percentage increase e.g. 5%, 10%.

Questions:

- What are Members opinions on the PROAST and BMDS software approaches?
- The differences in the two models have been described above; do these differences affect the outcome? And if so in what way?
- Is there justification to use one model rather than the other?
- Should models be constrained or not?

Davis JA, Gift JS, Zhao QJ. (2011) Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1.Toxicol Appl Pharmacol. 254(2):181-91.

EFSA (2009) Use of benchmark dose approach in risk assessment EFSA J., 1150 (2009), pp. 1–72

EFSA (2011) TECHNICAL REPORT: Use of BMDS and PROAST software packages by EFSA Scientific Panels and Units for applying the Benchmark Dose (BMD) approach in risk assessment. EN-113. [190 pp.]. Available online: www.efsa.europa.eu

http://onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2011.EN-113/pdf

EFSA (2016) Public Consultation on the draft update of the guidance of the Scientific Committee on the use of the benchmark dose approach in risk assessment.

https://www.efsa.europa.eu/sites/default/files/consultation/160714.pdf