This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

CC/2017/20

COMMITTEE ON CARCINOGENICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT (COC)

Update on benchmark dose modelling

Introduction

1. As part of the revisions to Guidance Statement G05 on points of departure and potency estimates, discussed in July 2017, the COC requested an overview provided here of recent updates to benchmark dose (BMD) modelling. In particular, the updated EFSA (2017) guidance will be considered.

Abbreviations

ADI	Acceptable daily intake
AIC	Akaike information criterion
BMD	Benchmark dose
BMDL	Lower 95% confidence limit of the benchmark dose
BMDS	Benchmark dose software
BMDU	Upper 95% confidence limit of the benchmark dose
BMR	Benchmark response
EFSA	European Food Safety Authority
HBGV	Health-based guidance value
IPCS	WHO International Programme on Chemical Safety
MOE	Margin of exposure
NOAEL	No observed adverse effect level
OECD	Organisation for Economic Co-operation and Development
PoD	Point of Departure
RP	Reference Point
WHO	World Health Organization

Literature search strategy

3. A search of the literature was performed by the National Centre for Environmental Toxicology at WRc (NCET at WRc) and IEH-Consulting Ltd. (IEH-C) (NCET at WRc/IEH-C) under contract to PHE on 01/09/2017, to identify guidance on the use of benchmark dose (BMD) modelling in risk assessment published between 2009 and 2017. Both the World Health Organization (WHO) / International

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

Programme on Chemical Safety (IPCS) and European Food Safety Authority (EFSA) have produced guidance on dose–response modelling, including guidance on cancer dose-response data (IPCS, 2009 and EFSA, 2009, 2017). The most recent update on the use of BMD modelling in risk assessment is that by EFSA which takes additional experience accumulated since the prior publication in 2009. A summary of the main points from the EFSA 2017 update are included below.

Use of dose-response data in hazard characterisation

4. Two possible approaches are currently available for deriving a Reference Point (RP) (point of departure) from animal toxicity studies, the No Observed Adverse Effect Level (NOAEL) approach and the BMD approach.

5. EFSA (2017) concludes that the BMD approach is a "*scientifically more advanced method compared to the NOAEL approach for deriving a RP"*, that can be applied to dose-response data for all toxicological effects from experimental and epidemiology studies. This is qualified by the extended use of dose–response data in the BMD approach that allows quantification of the uncertainty in the estimated RP. It is emphasised that use of the BMD approach does not remove the need for a critical evaluation of the response data by the assessor. In addition, it should be noted that current guidelines are optimised to identify the NOAEL and may not be optimal for deriving the BMD. EFSA (2017) states that as testing guidelines are revised, study designs that result in better dose-response information (more dose levels with the same number of animals) should be recommended.

6. EFSA (2017) reports that a more consistent RP is derived using the BMD approach, due to the specified benchmark response (BMR) (see paragraph 11), which has benefits for establishing health-based guidance values (HBGVs) or calculating margin of exposures (MOEs). HBGVs derived using the BMD approach are considered to be as protective as those derived using the NOAEL approach. Further, the default values for uncertainty factors currently applied to NOAEL values are equally applicable to the BMD.

The BMD approach

7. The BMD is defined by EFSA as '*a dose level, estimated from the fitted dose–response curve, associated with a specified change in response, the BMR*'. The BMDL is the BMD's lower confidence bound, and this value is normally used as the RP. The key concepts in the BMD approach are illustrated in Figure 1 and its legend, taken from the EFSA (2017) guidance.

8. Figure 1 shows that a BMDL calculated for example for a BMR of 5%, can be interpreted as follows:

$BMDL_{05}$ = dose where the change in response is likely to be smaller than 5% where the term 'likely' is defined by the statistical confidence level, usually 95% confidence.

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

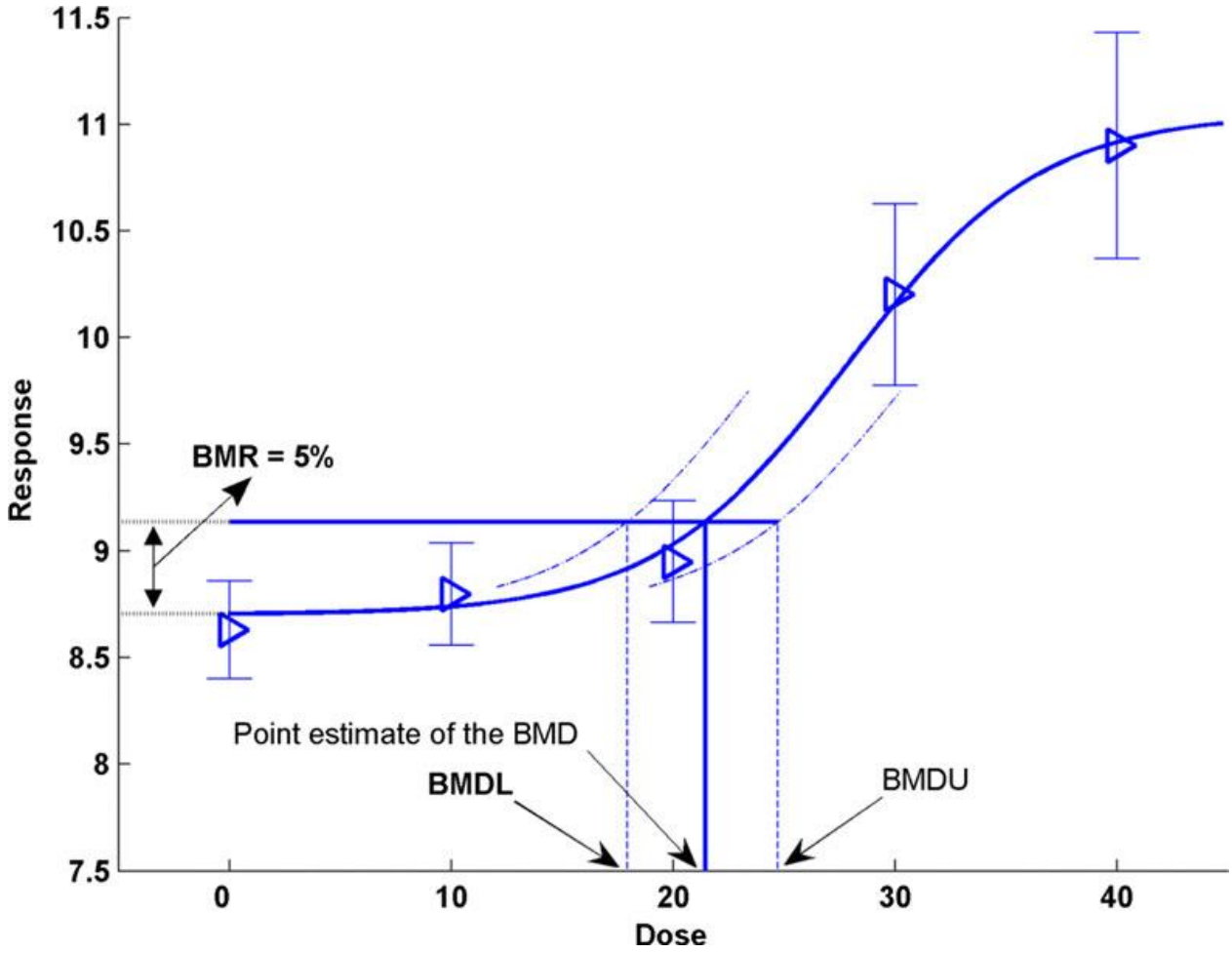

Figure 1: Key concepts for the BMD approach, illustrated using hypothetical continuous data.

Source EFSA (2017). The observed mean responses (triangles) are plotted, together with their confidence intervals. The solid curve is a fitted dose–response model. This curve determines the point estimate of the BMD, which is generally defined as a dose that corresponds to a low but measurable change in response, denoted the benchmark response (BMR). The dashed curves represent, respectively, the upper and lower 95% confidence bounds (one sided) for the effect size as a function of dose. Their intersections with the horizontal line are at the lower and upper bounds of the BMD, denoted BMDL and BMDU, respectively. It should be noted that the BMR is not defined as a change with regard to the observed mean background response, but with regard to the background response predicted by the fitted model. This distinction is important because, in general, the fitted curve does not hit the observed background response exactly (so that adding the BMR to the observed background response will in general not provide the correct intersection with the dose–response at the BMD). In the Figure, the BMD corresponds to a 5% change in response relative to background (BMR = 5%). The fitted curve yields an estimated background response of 8.7, and a 5% increase of that equals 9.14 (= 8.7 + 0.05 x 8.7). Thus, the BMD_{05} of 21.50 is obtained from the intersection of the horizontal line, at a response of 9.14, with the fitted dose–response model. In this example, the $BMDL_{05}$ has a value of 18.

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

9. The application of the BMD approach may be summarised as a process involving the following steps:

i. Specification of type of dose–response data.
ii. Specification of the BMR.
iii. Selection of candidate dose–response model(s).
iv. Fitting the candidate models and calculate the BMD confidence interval for each model.
v. Combining the results from the various models into one single BMD confidence interval, with the lower bound (BMDL) as the RP.

10. ***Specification of dose-response data type***: response data may be of various types including, most commonly, continuous and quantal. Identification of data type is important with BMD modelling for selecting mathematical and statistical models and interpretation of the BMR. For continuous data (e.g. liver weight; enzyme activity) the BMR is a measure of the degree of severity of effect, whereas for quantal data (e.g. presence or absence of abnormal liver status) the BMR reflects a change in incidence. Other, less common, types of biological data include ordinal type (e.g. histology scores ranging from 1-normal to 5-extremely abnormal).

11. ***Specification of BMR***: this is a specific value of the effect that is used to define the BMD. Several options are available depending on whether quantal or continuous data is being considered. With regards to quantal data, the BMR is defined in terms of an increase in the incidence of the response scored, compared with the background incidence. For human studies, commonly used metrics include additional risk and relative risk. Ideally the BMR should reflect a negligible or non-adverse effect size. BMR values for experimental data of 1, 5, and 10% (extra or additional risk) were initially proposed (Crump, 1984; EPA, 1995), however EFSA (2017) suggest a default BMR of 10% should be adopted for quantal data since the modelling of lower responses generally results in greater uncertainty. The $BMDL_{10}$ has been reported to be on average close to the NOAELs for lethality and developmental toxicity data. For continuous data, several definitions for the BMR are possible, however a default of 5% is suggested by EFSA (2017). Ideally the BMR should be expressed as a per cent change in mean response as compared to the background response; this defines a BMR that is independent of within-group variation. A re-analysis of a large number of National Toxicology Program (NTP) studies (Bokkers and Slob, 2007) showed that the $BMDL_{05}$ was, on average, close to the NOAEL derived from the same dataset.

12. ***Selection of dose-response models***: the BMD approach uses simple statistical models that do not describe the underlying biology. This is because BMD analysis aims to find all plausible values of the (true) BMD, given the data available, which is achieved by assessing both the best-fitting and poorer fitting models. This ensures that the BMD confidence intervals are based on the results from various models, instead of just a single (‘best’) model (EFSA, 2017).

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

13. Currently available models for quantal and continuous data are summarised in Table 1, including two models that relate to both types of data, the full (or saturated) model and the null model. The full-model does not assume any specific dose-response and can be used to evaluate the goodness of fit of any dose-response model. The null model assumes no dose-related trend is present and can be used to statistically evaluate the presence of a dose-related trend (EFSA, 2017).

Table 1: Recommended models for use in the BMD approach (source EFSA, 2017)

Model	**Number of model parameters**	**Model expression mean response (*y*) as function of dose (*x*)**	**Constraints**
Full model[(i)]	Number of dose groups including background	Set of observed means or incidences at each dose	ns
Null model[(ii)]	1	y = a	a> 0 for continuous data 0 < a < 1 for quantal data
Continuous data			
Exponential family			
3-parameter model[(iii)]	3	$y = a \exp(bx^d)$	a > 0, d > 1
4-parameter model[(iv)]	4	$y = a [c-(c-1)\exp(-bx^d)]$	a > 0, b > 0, c > 0, d > 0
Hill family			
3-parameter model[(iii)]	3	$y = a [1\_xd/(bd + x^d)]$	a > 0, d > 1
-4-parameter model[(iv)]	4	$y = a [1 + (c-1)x^d/(b^d + x^d)]$	a > 0, b > 0, c > 0, d > 0
Quantal data			
Logistic	2	y = 1 / (1 + exp(-a – bx))	b > 0
Probit	2	y = CumNorm(a + bx)	b > 0
Log-logistic	3	y = a + (1-a) / (1 + exp(-log(x/b)/c))	$0 \leq a \leq 1$, b>0, c>1
Log-probit	3	y = a + (1-a) CumNorm(log(x/b)/c)	$0 \leq a \leq 1$, b>0, c>0
Weibull	3	$y = a + (1-a) \exp((x/b)^c)$	$0 \leq a \leq 1$, b>0, c>1
Gamma	3	$y = a + (1-a)$ CumGam(bx^c)	$0 \leq a \leq 1$, b>0, c>1
LMS (two-stage) model	3	$y = a + (1-a) \exp(– bx – cx^2)$	a>0, b>0, c>0
Latent variable models based on CM [(v)]	Depends on underlying CM	Assume an underlying continuous response dichotomised into yes/no response based on a (latent) cut-off value that is estimated from the data	As for continuous models
a, b, c, d: unknown parameters that are estimated by fitting the model to the data. **CumNorm**: cumulative (standard) normal distribution function. **CumGam**: cumulative Gamma distribution function			

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

ns – not specified by EFSA; **CM** – continuous model
(i): The full model will result in the maximum possible value of the log-likelihood (given the statistical assumptions) for the data set considered.
(ii): The null model can be regarded as a model that is nested within any dose–response model: it reflects the situation of no dose response (= horizontal line).
(iii): Called model 3 in PROAST, and similarly (for the exponential model) in BMDS.
(iv): Called model 3 in PROAST, and similarly (for the exponential model) in BMDS.
(v): The latent variable models are implemented in PROAST.

14. The dose-response models also need to describe the within-(dose)-group variation, which is referred to as the 'distributional part' of the dose-response model (EFSA, 2017). For continuous data, this variability is seen as the scatter of individual data points around the fitted curve and described by a statistical distribution. The log-normal distribution is proposed as the default choice by EFSA, giving the mean and standard deviation of the response at a stated dose; geometric mean and standard deviations can also be derived (Slob, 1994). Due to the difficulties in identifying dose-dependent within-group variation, EFSA also recommends that the default assumption should be constant variability among dose groups (EFSA, 2017).

15. The within-group variation for quantal data is normally not directly visible. However, this observed incidence is subject to random sampling error which is binomially distributed (when no dependencies exist); this is the default assumption for quantal data. Where dependencies are identified, such as litter effects, these can be taken into account in the modelling software (see paragraphs 17-18).

16. The currently available BMD software packages perform 'best fitting' for each model, to find values for unknown parameters that approach the known data as closely as possible. This is achieved by maximising the log-likelihood that can be reached by that model. Fitting is assessed through evaluating the following:

i. Convergence of fitting algorithms within the BMD software which occurs when the log-likelihood can no longer be achieved. If convergence is not reached this may indicate that the data did not provide sufficient information to appropriately estimate all the parameters in the model.
ii. The Akaike information criterion (AIC), which directly integrates the log-likelihood and the number of model parameters into a single value; the lower the value determined for a model, the better the fit. Models differing in AIC value by less than two units are considered to describe the data equally well.
iii. Covariate analysis, which is carried out when fitting models to a combination of data sets differing by a specific aspect (e.g. species) to assess any differences in dose-responses between sub-groups, or to obtain a smaller confidence interval for a BMD.

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

BMD software

17. Different software programs are currently available for BMD analysis. The US Environmental Protection Agency (US EPA) developed the Benchmark Dose Software (BMDS) which currently contains thirty different models. PROAST is an additional BMD software package, developed by the Dutch National Institute for Public Health and the Environment (RIVM), and is the basis on which EFSA now provides a web-based platform for performing BMD analysis. Both these are suitable for dose-response analysis and deriving a BMDL from the dose-response data. RIVM and US EPA collaborate to achieve consistency between the BMDS and PROAST software. A summary of the main differences between the BMDS and PROAST software is provided in Table 2.

18. BMDS and PROAST differ in their default assumptions regarding the distribution of continuous data. In BMDS data are assumed to be normally distributed, whilst in PROAST a log-normal distribution is assumed. EFSA (2017) highlights the lack of option within the BMDS software to set the distribution to log-normal in the Hill model and that the three-parameter Hill model cannot be fitted. Although the parameters within the Hill model differ between the two software packages, they can be regarded as equivalent. The BMDS software also contains additional models to PROAST, including the four-parameter Hill model, the nested family of exponential models, and power, linear and polynomial models for continuous data.

Table 2: Comparison of BMDS and PROAST

	BMDS	**PROAST**
Environment	Can be run immediately (as an executable) under Windows	R (free software) is required Also runs under Linux and Mac OS X
First use	Easy to get started	Higher threshold; requires basic understanding of R
User interaction	Graphical User Interface (GUI)	Both a menu version and a GUI version available. GUI is suitable for most standard analyses; the menu version covers more options
Continuous data	Yes	Yes
Nested continuous data, e.g. for litter effects	No	Yes
Quantal data	Yes	Yes (in menu version)
Nested quantal data, e.g. for litter effects	Yes	Yes
Ordinal data	A program on categorical regression is implemented	Yes

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

	BMDS	**PROAST**
BMDU calculated	No, except for Multistage Cancer model	Yes
Default assumption of distribution continuous data	Normal	log-normal
Option to change default distribution continuous data	Only for exponential model	Yes (in menu version)
Confidence interval based on profile likelihood	Yes	Yes
Confidence interval based on bootstrapping	No	Yes (in menu version)
Covariates	No (except for nested quantal models)	Yes
Model fitting for (nested) exponential models	Yes	Yes
Model fitting for (nested) Hill models	No, only four-parameter model	Yes
Automatic model fitting for recommended suite of quantal models	Yes	Yes
Graphical output	Yes, but only original scales for y-axis and x-axis	Yes, including options to change scales (e.g. log-scales)
Evaluation of dose addition	No	Yes

Model averaging and establishing the BMD confidence interval

19. The BMD approach finds all relevant values that are compatible with the data and not a single statistically best estimate, and therefore the BMD value should take into account results from all models applied. Whilst acknowledging that respective tools are still under development, EFSA recommends using the Model Averaging approach as the preferred method for calculating the confidence interval for a BMD, which is present in the PROAST software (EFSA, 2017). Model averaging accounts for both model uncertainty and for the uncertainty related to sampling errors in the data (Burnham and Anderson, 2004; Wheeler and Bailer, 2007, 2008, 2009).

20. The Model Averaging approach combines the model output using a weighting of models, with a higher weight given for better fitting models; weights are often defined in terms of the AIC. In the first step, the average response is calculated for a large number of doses, by taking the weighted average from the fitted dose–response models involved and the BMD calculated for the average model. In the second step, a large number of artificial data sets are generated based on the average model, and for each these the first step is repeated. The process generates a large number of BMDs, the lower and higher 5th percentiles of these then define the BMD 90% confidence interval.

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

21. Where model averaging tools are unavailable, models can be selected and rejected on a 'goodness of fit' basis based on the AIC value which should generally be no more than two units bigger than the minimum AIC (i.e. the best fitting model) (Burnham and Anderson, 2004). The lowest BMDL and highest BMDU from these selected models are used to define the BMD confidence interval.

22. Once the models have been fitted, data are assessed as shown in Figure 2. The process breaks down into 3 main steps:

Step 1: When the software reports 'no convergence' for one or more models, this is an alert that the data are not very informative, or the model is over-parameterised.

Step 2: Checks for dose-related trends from at least one model. Dose related trends are defined when an AIC is lower than the AIC of the null model + 2 units. For continuous data, the model resulting in the lowest AIC is selected.

Step 3: Results from the fitted models are combined to establish the final BMD confidence interval, preferably using the model averaging approach.

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

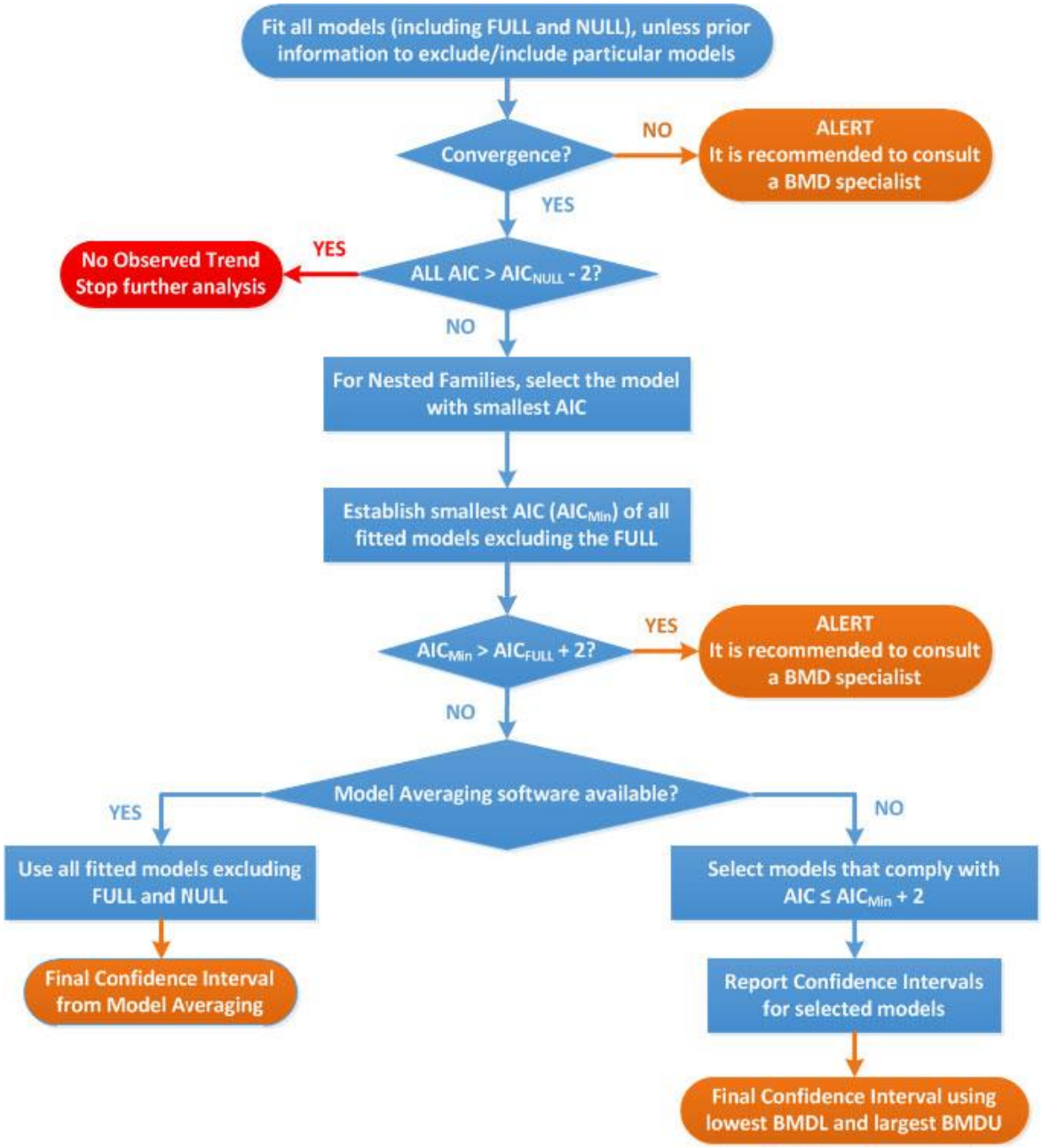

Figure 2: Flow chart to establish the BMD confidence interval and BMDL for dose-response data set of a specified endpoint.

AIC: Akaike information criterion (indicative of the goodness of fit of the model considered); **AICnull**: AIC value of the Null Model; **AICfull**: AIC value of the Full Model; **AICmin**: AIC value of the model with the lowest AIC value, the null and full models being excluded. Source; EFSA (2017).

Quality criteria

23. The acceptability of a dataset to derive a NOAEL as a potential RP depends on expert judgement, as poor or limited data tend to result in high NOAELs. In addition to deriving a RP, the BMD approach also quantitatively evaluates data quality. There are two situations that indicate that the data are not informative enough to derive an RP: (i) when the BMDL/BMDU ratios for the individual models are very large, including situations where the BMDL may approach 'zero', e.g. with

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

data where the lowest dose tested results in a response much higher than the BMR; and (ii) when the BMDLs among models are very different (in data with high model uncertainty).

Consequences of the use of BMD for hazard / risk characterisation (impact on COC Guidance doc)

24. The use of a RP derived using BMD modelling (e.g. BMDL) in risk assessment does not change the basic approach or assumptions. EFSA (2017) strongly recommends that the BMD approach, and more specifically model averaging, is used for the determination of RPs for use in: (i) deriving HBGVs; (ii) MOE approach for substances that are both genotoxic and carcinogenic; (iii) comparison of potencies; and (iv) probabilistic risk assessment.

Question for the Committee

i. Is there newly available information on BMD which should be incorporated in the revision of G05?

NCET at WRc/IEH-C under contract supporting the PHE COT Secretariat October 2017

This is a preliminary paper for discussion. It does not represent the views of the Committee and must not be quoted, cited or reproduced.

References

Bokkers BGH and Slob W (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the Benchmark dose approach. Crit. Rev. Toxicol. 37: 353-377.

Burnham KP and Anderson DR (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research, 33, 261–304.

Crump KS (1984). An improved procedure for low-dose carcinogenic risk assessment from animal data. J Environ Pathol Toxicol Oncol. 5: 339–348.

EFSA (2009). Guidance of the Scientific Committee on use of the benchmark dose approach in risk assessment. The EFSA Journal 1150, 40-47

EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen KH, More S, Mortensen A, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Silano V, Solecki R, Turck D, Aerts M, Bodin L, Davis A, Edler L, Gundert-Remy U, Sand S, Slob W, Bottex B, Abrahantes JC, Marques DC, Kass G and Schlatter JR (2017). Update: Guidance on the use of the benchmark dose approach in risk assessment. EFSA Journal 2017;15(1):4658, 41 pp. doi:10.2903/j.efsa.2017.4658

IPCS (2009). Principles for Modelling Dose-Response for the Risk Assessment of Chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria 239).

Slob W (1994). Uncertainty Analysis in Multiplicative Models. Risk Analysis, 14, 571–576.US EPA, 1995. The use of the benchmark dose approach in health risk assessment. EPA/630/R-94/007.Risk Assessment Forum, Washington DC.

Wheeler MW and Bailer AJ (2007). Properties of Model-Averaged BMDLs: A Study of Model Averaging in Dichotomous Response Risk Estimation. Risk Analysis, 27, 659–670.

Wheeler MW and Bailer AJ (2008). Model averaging software for dichotomous dose response risk estimation. Journal of Statistical Software, 26, 1–15.

Wheeler MW and Bailer AJ (2009). Comparing model averaging with other model selection strategies for benchmark dose estimation. Environmental and Ecological Statistics, 16, 37–51.