GUIDANCE DOCUMENT

EFSA Journal

ADOPTED: 5 December 2014 PUBLISHED: dd mmmm yyyy AMENDED: dd mmmm yyyy

doi:10.2903/j.efsa.20YY.NNNN

Use of the benchmark dose approach in risk assessment

EFSA Scientific Committee

Abstract

(Max. 300 words, no paragraph breaks; no tables, footnotes, graphs or figures. Note that the abstract should end with the copyright)

Keywords: Benchmark dose, BMD, BMDL, benchmark response, NOAEL, dose-response modelling, BMD software

Requestor: EFSA

Question number: EFSA-Q-2014-00747

Correspondence: sc.secretariat@efsa.europa.eu

Scientific Committee members: Diane Benford, Thorhallur Halldorsson, Anthony Hardy, Michael John Jeger, Katrine Helle Knutsen, Simon More, Alicja Mortensen, Hanspeter Naegeli, Hubert Noteborn, Colin Ockleford, Antonia Ricci, Guido Rychen, Josef R. Schlatter, Vittorio Silano, Roland Solecki and Dominique Turck.

Acknowledgements: The Panel wishes to thank the members of the Working Group on Benchmark Dose: Marc Aerts, Laurent Bodin, Allen Davis, Lutz Edler, Ursula Gundert-Remy, Salomon Sand, Josef R. Schlatter (Chair) and Wout Slob for the preparatory work on this scientific output, and EFSA staff members: Bernard Bottex, Jose Cortinas Abrahantes, Daniele Court Marques and Georges Kass for the support provided to this scientific output.

Suggested citation: EFSA Scientific Committee, 2016. Draft update of the guidance of the Scientific Committee on the use of the benchmark dose approach in risk assessment. EFSA Journal 20YY;volume(issue):NNNN, 51 pp. doi:10.2903/j.efsa.20YY.NNNN

ISSN: 1831-4732

The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.

Summary

Considering the need for transparent and scientifically justifiable approaches to be used when risks are assessed by the Scientific Committee and the Scientific Panels of EFSA, the Scientific Committee was requested in 2005 by EFSA i) to assess the existing information on the utility of the benchmark dose (BMD) approach, as an alternative to the traditionally used NOAEL approach, ii) to provide guidance on how to use the BMD approach for analysing dose-response data from experimental animal studies, and iii) to look at the possible application of this approach to data from observational epidemiological studies.

The Scientific Committee confirmed that the BMD approach is a scientifically more advanced method compared to the NOAEL approach for deriving a Reference Point (RP), since it makes extended use of available dose-response data and it provides a quantification of the uncertainties in the dose-response data. The BMD approach is therefore recommended as the approach to be used by EFSA's Panels and Units to derive RP's for establishing health based guidance values or for calculating margins of exposure between the RP and the human exposure.

A guidance document on the use of the benchmark dose approach in risk assessment was published in 2009 (EFSA, 2009). In 2015, the Scientific Committee reviewed the implementation of the BMD approach in EFSA's work, the experience gained with its application, and the latest methodological developments in regulatory risk assessment, and concluded that an update of its guidance from 2009 was necessary. Most of the modifications made to the SC guidance of 2009 concern the section providing guidance on how to apply the BMD approach in practice (section 2.5). 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.

The Scientific Committee reconfirms in this updated guidance that the BMD approach, and more specifically model averaging, should be used for deriving a RP from the critical dose-response data to establish health-based guidance values and margins of exposure. This updated guidance does not require reconsidering previous assessments where the NOAEL approach was used, or where quantal datasets were analysed using the BMD approach as described in the 2009 SC guidance. The application of this updated guidance to previous risk assessments where the 2009 guidance was used to analyse continuous datasets might result in lower RP's, in particular when model 2 of the nested families was selected to derive the RP.

The Scientific Committee recommends that training in dose-response modelling and the use of BMD software continues to be offered to experts in the scientific Panels and EFSA Units. EFSA should establish a Standing Working Group on BMD analysis to be consulted by EFSA experts and staff if needed, e.g. when alerts are identified or when applying the BMD approach to histopathological (ordinal) data. A network on BMD, coordinated by EFSA, should also be considered to exchange experience and develop expertise with EFSA Partners (Member States competent, EU sister agencies, DG Santé Scientific Committees and international organisations).

The Scientific Committee also identified the need for a specific guidance on the use of the BMD approach to analyse human data.

Finally, the Scientific Committee firmly reiterated the need for current toxicity test guidelines to be reconsidered given the expected wide application of the BMD approach.

Table of contents

1. Introduction

As per EFSA's Founding Regulation (EC) No 178/2002 of the European Parliament and of the Council, "the EFSA Scientific Committee shall be responsible for the general coordination necessary to ensure the consistency of the scientific opinion procedure, in particular with regard to the adoption of working procedures and harmonisation of working methods". The EFSA Science Strategy 2012-2016 echoes this key responsibility of the Scientific Committee by setting the development and harmonisation of methodologies and approaches to assess risks associated with the food chain as one of the four strategic objectives for EFSA.

In May 2009, the Scientific Committee adopted its guidance on the use of the benchmark dose (BMD) approach in risk assessment (EFSA Scientific Committee, 2009). When deriving a reference point (point of departure), the guidance document recommends using the BMD approach instead of the traditionally used NOAEL approach, since it makes a more extended use of dose-response data and it allows for a quantification of the uncertainties in the dose-response data. The BMD approach is applicable to all chemicals in food, irrespective of their category or origin.

Feedback was gathered by EFSA's Secretariat regarding the implementation of this approach by EFSA's Scientific Panels during the last 7 years; several issues were highlighted as worth further clarification. During its 67th Plenary meeting (see minutes), the Scientific Committee agreed with the proposal to update the guidance document on the use of the benchmark dose approach in risk assessment.

1.1. Terms of Reference as provided by EFSA

The European Food Safety Authority requests the Scientific Committee to update the existing guidance to clarify the following issues:

1. The current SC guidance recommends specific mathematical models to be fit for quantal and continuous dose-response data. The list of recommended models should be reviewed.
2. The approach for selecting the best model when dealing with a family of nested models should be reviewed; suggestion was made to deviate from the approach recommended in the SC guidance and use directly the full versions of the nested family of models, i.e. the Exponential and Hill four parameters models when dealing with continuous data (Slob and Setzer, 2014).
3. The need and relevance to constrain some of the model parameters so that they reflect the biology should be reviewed and guidance should be provided on when / when not to constrain models.
4. Further guidance should be provided on how to deal with poor datasets. Criteria regarding acceptable BMDL[1]-BMD or BMDL-BMDU[2] intervals to derive a reference point should be provided, as well as recommendations on how to deal with datasets that do not comply with these criteria.
5. Further guidance should be provided on using covariate analysis and when combining data for a BMD analysis.

This list is non-exhaustive and may be expanded with additional issues identified during the updating process, if deemed worth further clarification.

[1] BMDL: lower bound of the BMD confidence interval
[2] BMDU: upper bound of the BMD confidence interval

2. Assessment

2.1. Introduction

This opinion primarily addresses the analysis of dose-response data from experimental studies but also considers the application to data from observational epidemiological studies. Similar approaches can also be applied to ecotoxicity studies but are not further considered in this guidance. Toxicity studies are conducted to identify and characterize the potential adverse effects of a substance. The data obtained in these studies may be further analysed to identify a dose that can be used as a starting point for risk assessment. The dose used for this purpose, however derived, is referred to in this opinion as the Reference Point (RP). This term has been used already by EFSA in the opinion of the Scientific Committee on a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic (EFSA 2005), and is therefore preferred to the equivalent term Point of Departure (PoD), used by others such as US EPA.

The No-Observed-Adverse-Effect-Level (NOAEL) has been used historically as the RP for estimating health-based guidance values such as acceptable daily intakes (ADIs), tolerable daily intakes (TDIs), or tolerable weekly intakes (TWIs) in risk assessment of non-genotoxic substances.

EFSA (2005) and the Joint FAO/WHO Expert Committee on Food Additives (JECFA, 2006a) have proposed the use of the BMD approach for deriving the RP used to calculate the margin of exposure for substances that are both genotoxic and carcinogenic. As the NOAEL is known to have some limitations (see following sections), the Scientific Committee concluded in 2009 that the Benchmark Dose (BMD) approach is the best approach for defining a RP also for non-genotoxic substances (EFSA, 2009). The methodology discussed in this guidance document has subsequently been applied for deriving RPs (i.e. BMDLs) for various types of chemicals (e.g. pesticide, additives and contaminants). The Scientific Committee reviewed in 2015 the implementation and the experience of the BMD approach in EFSA's work, as well as the latest methodological developments in regulatory risk assessment to prepare the present update of its guidance document.

In sections 2.1 to 2.3 of this guidance document, the concepts underlying both the NOAEL and BMD approaches are discussed with some illustrative examples. In these sections, it is outlined why the Scientific Committee considers the BMD approach as the more powerful approach. Section 2.4 discusses the potential impact of using the BMD approach for hazard/risk characterisation and risk communication. Section 2.5, which provides guidance on how to apply the BMD approach in practice, has been significantly modified compared to the 2009 version of the guidance document: model averaging is more strongly emphasized as the preferred method for calculating the BMD confidence interval. Further, the set of default models to be used for BMD analysis has been revised while the evaluation of model performance is now based on the so-called AIC (Akaike information criterion) instead of the log-likelihood. At the end of section 2.5, two examples, one based on quantal data, the other on continuous data, are provided to illustrate the application of the BMD approach in practice and how to report the results. A template for BMD analysis reporting has been inserted in Appendix B.

The primary audience for this guidance document comprises all those contributing to EFSA's scientific assessments. The Scientific Committee considers that the use of the BMD approach is always better than the NOAEL approach to define a RP; therefore the application of this guidance document is unconditional for EFSA and is strongly recommended for all parties submitting assessments to EFSA for peer-review.

2.2. Hazard identification: selection of potential critical endpoints

Toxicity studies are designed to identify the adverse effects produced by a substance, and to characterize the dose-response relationships for the adverse effects detected. While in some cases human dose-response data are available, most risk assessments rely on data from animal studies. The aim of hazard identification is to identify potential critical endpoints that may be of relevance for human health. An important component is the consideration of dose dependency of observed effects. Traditionally this is done by visual inspection, together with conventional statistical tools. The Scientific Committee recommends using dose-response modelling approaches instead.

Selection of the critical effect should not be based on the statistical procedures only. Importantly, additional toxicological arguments should be taken into account in the evaluation of a full toxicological data package. Use of the BMD approach does not remove the need for a critical evaluation of the response data[3] and an assessment of the relevance of the effect to human health.

The result of this first step is the identification of potential critical endpoints that should be analysed in more detail as described in the next sections.

2.3. Using dose-response data in hazard characterisation

The nature of the dose-response relationships is explored in detail in hazard characterisation. For most toxicological effects, the overall aim of the process is to identify a dose without appreciable adverse health effects in the test animals under the experimental conditions. The RP from the toxicity studies is then used to establish a level of human intake at which it is confidently expected that there would be no appreciable adverse health effects, taking into account uncertainty and variability such as inter- and intra-species differences, suboptimal study characteristics, missing data.

Hazard characterisation in risk assessment requires the use of a range of dose levels in animal toxicity studies. Doses are needed that produce different effects sizes providing information on both the lower and higher part of the dose-response relationship to characterise the full dose-response relationship.

Experimental and biological variations affect response measurements; in consequence, the mean response at each dose level will include a statistical error. Therefore, dose-response data need to be analysed by statistical methods to prevent inappropriate biological conclusions being drawn because of statistical errors associated with the data. Currently, there are two statistical approaches available for deriving a reference point: the NOAEL approach, and the BMD approach. This section reviews these two approaches, and provides a comparison of the strengths and limitations of each method.

2.3.1. The NOAEL approach

The NOAEL approach is applicable to all toxicological effects considered to have a threshold.

The study NOAEL is derived as follows:

- For each adverse effect/endpoint, identify the highest experimental dose level where effects were not detected, using expert opinion and statistical tests to compare each dose group with the control group.
- The study NOAEL is the lowest relevant NOAEL obtained for any of the adverse effects detected in the study (i.e. for the critical effect of the study).

Hence, the NOAEL is the highest dose tested without evidence of an adverse effect in the particular experiment. The numerical value of the NOAEL is thus dependent upon the selection of dose levels when the study was designed and on the ability of the study to detect adverse effects. Since studies with low power (e.g. small group sizes) and/or insensitive methods are able to detect only relatively large effects, these tend to result in higher NOAELs. If there is a statistically significant effect at all dose levels, the lowest dose used in the study is usually selected as the lowest-observed-adverse-effect-level (LOAEL).

It should be noted that in general, identification of the NOAEL is not always a purely statistically-based decision. This has advantages and disadvantages. It allows the assessor to reject a NOAEL which is not well supported by the data, but it can also lead to different decisions. Factors that may be taken into account in identification of the NOAEL include whether there is a consistent dose-response relationship and the magnitude of the change in (mean) response. When the observed change in response is small, even if statistically significant, some assessors may consider it non-adverse and use a higher dose as the NOAEL. In contrast, where there is a small, non-significant increased response in the dose group below the statistically significant effect, some assessors may

[3] In this opinion, "response" is used as a generic term that refers to both quantal and continuous data.

identify this response still as an adverse effect (i.e. being a LOAEL). Such decisions are based on expert judgement and different assessors may reach different decisions, as happened in the past, e.g. in the evaluation of residues of the veterinary drug ractopamine by JECFA (2006b) and the EFSA Panel on additives and products or substances used in animal feed [FEEDAP] (EFSA, 2009a).

2.3.2. The BMD approach

The BMD approach is applicable to all toxicological effects. It makes use of all of the dose-response data to estimate the shape of the overall dose-response relationship for a particular endpoint. The BMD is a dose level, estimated from the fitted dose-response curve, associated with a specified change in response, the Benchmark Response (BMR), (see Section 2.5.2). 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 Fig. 1 and its legend. This figure shows that a BMDL that is calculated, e.g. 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.

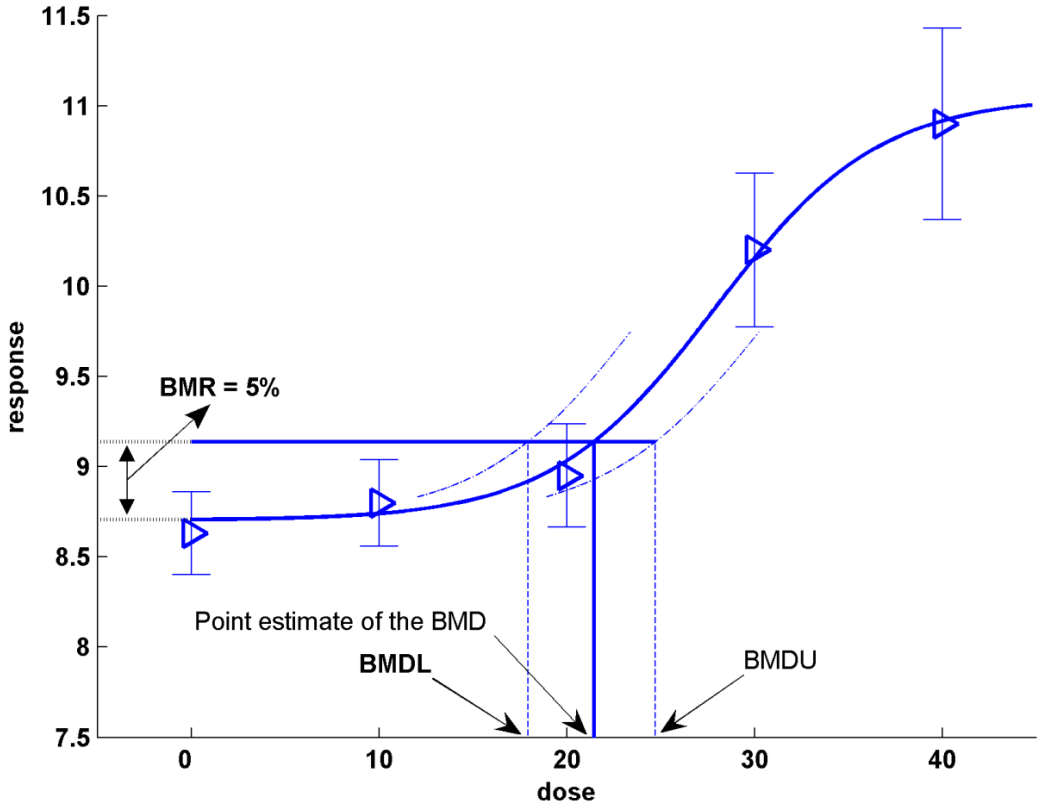


Figure 1: Key concepts for the BMD approach, illustrated by using hypothetical continuous data.

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)[4] 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.

[4] A lower (or upper) 95%-confidence bound (one-sided) is equivalent to the lower (or upper) limit of a two-sided 90%-confidence interval.

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*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.

The essential steps involved in identifying the BMDL for a particular study are:

- Specification of a response level, e.g. a 5% or 10% increase or decrease in response compared with the background response. This is called the BMR (see section 2.5.2).
- Fitting a set of dose-response models (section 2.5.3), and calculation of the BMD confidence interval for each of the models that describe the data according to statistical criteria, resulting in a set of BMD confidence intervals.
- Deriving a single BMD confidence interval from the set of BMD confidence intervals for that particular adverse effect/endpoint, preferably by model averaging (section 2.5.6).
- An overall study BMDL, i.e. the critical BMDL of the study, is selected from the obtained set of BMD confidence intervals for the different potentially critical endpoints (see section 2.5.7).

In principle, the BMD approach could be applied to every endpoint measured in the relevant studies. The critical effect would then be selected in an analogous way as in the NOAEL approach, that is, as the endpoint resulting in the lowest BMDL, but also taking additional toxicological arguments into account, just as in the case of the NOAEL approach. However, it is recommended to make use of one of the strengths of the BMD approach, and select the study BMDL based on considering the complete BMD confidence intervals for the endpoints considered and combine the information on uncertainties in the underlying data with biological considerations (see section 2.5.7). In the NOAEL approach the decision to accept a dataset for deriving a NOAEL as a potential RP is important since poor or limited data (e.g. due to high variability within the dose groups, high limit of quantification of analytical methods, small sample sizes) will tend to result in high NOAELs. Acceptability of the data will therefore depend upon expert judgement. In contrast, the BMD approach itself provides a formal quantitative evaluation of data quality, by taking into account all aspects of the specific data. When the data are relatively poor or uninformative, the resulting BMD confidence interval for that dataset will tend to be wide, and the BMDL might be much lower than the true BMD. But the meaning of the BMDL value remains as it was defined: it reflects a dose level where the associated effect size is unlikely to be larger than the BMR used.

Nonetheless, it might happen that the data are so poor that using the associated BMDL as a potential RP appears unwarranted, and the dataset may need to be discarded. This might be decided when the confidence intervals around the BMD are wide or when different models result in widely different BMDL values. This issue is further discussed in section 2.5.7.

The most well-known BMD software are the benchmark dose software (BMDS) developed by the U.S. EPA (www.epa.gov/ncea), and the PROAST software developed by RIVM (www.rivm.nl/proast). Both software packages differ in some details as summarised in Appendix A.

2.3.3. Interpretation and properties of the NOAEL and the BMDL

The NOAEL is a dose level where generally no statistically significant differences in response are observed, compared with the background response. This implies that the NOAEL could reflect a dose level where effects are too small to be detected in that particular study, and therefore the size of the possible effect at the NOAEL remains unknown. A straightforward way of gaining insight into this is by calculating a confidence interval for the observed change in response between the control group and the NOAEL dose group.

For a limited number of substances, the Scientific Committee determined upper bounds for the effect size that are summarized in table 1. Here, the size of effect for quantal responses is expressed as extra risk. Extra risk is defined as an absolute change in frequency of response (additional risk in %) divided by the non-affected fraction in the control population (100 minus the background response in %)[5]. For continuous responses the effect size is expressed as a percent change in mean response[6]. For quantal endpoints, the upper bounds (which relate to extra risk) vary between around 3 and 30%. This illustrates that in some cases the extra risk at the NOAEL could be greater than 10%, which is the recommended BMR level for quantal data (see Section 2.5.2). Similarly, for this limited number of substances, it is found that the upper bound of the effect size at the NOAEL for continuous endpoints could be as small as 3%; but more often it was in the order of 10%, which is high compared with the 5% recommended for the BMR for continuous data (see section 2.5.2). In one of the examples, with a highly variable clinical-chemistry parameter, the upper bound of effect was as high as 260%.

The NOAEL is therefore not necessarily a "no adverse effect" dose, although it is sometimes interpreted as such. Indeed, as the review studies discussed in section 2.5.2 show, the size of the effect at the NOAEL is, on average over a number of studies, close to 10% (quantal responses) or 5% (continuous responses). For an individual NOAEL the size of effect remaining statistically non-significant might be smaller, or greater than these values. As illustrated in Table 1, it is possible to calculate an upper bound for the effect size at the NOAEL. Similarly, Sand et al. (2011) estimated that the median of the upper bounds of extra risk at the NOAEL was close to 10% based on analysis of about 800 datasets from the U.S. National Toxicology Program cancer bioassay database. However, the confidence interval of the effect size at the NOAEL is generally not reported in current applications. In the BMD approach, the potential size of the effect (i.e. the benchmark response, BMR) is by definition known.

For human (epidemiological) data, lower BMR values may be used because the observed response is often lower than 10% (see section 2.5.2).

[5] For example, when the additional risk is 8.5% and the background response is 15%, then the extra risk is 8.5/(100 – 15) = 10%.

[6] For example, a decrease in average BW from 100 g to 95 g would make an effect size of 5%.

327 **Table 1:** Illustrations of upper bounds[(a)] of effect at NOAELs related to 10 substances evaluated previously by JMPR or EFSA.

Substance (source +year)	Endpoint	Quantal data	Continuous data	References
		Upper bound extra risk (%)[(b)]	Upper bound percent change (%)[(c)]	
Thiodicarb (JMPR 2000)	splenic extramedullary haematopoiesis	21		www.inchem.org/documents/jmpr/jmpmono/v00pr09.htm
Carbaryl (JMPR 2001)	vascular tumours	15		www.inchem.org/documents/jmpr/jmpmono/2001pr02.htm
Spinosad (JMPR 2001)	thyroid epithelial cell vacuolation	2.7		www.inchem.org/documents/jmpr/jmpmono/2001pr12.htm
Flutolanil (JMPR 2002)	erythrocyte volume fraction haemoglobin concentration mean corposcular haemoglobin decreased cellular elements in the spleen	 30	9 9.7 3	www.inchem.org/documents/jmpr/jmpmono/2002pr07.htm
Metalaxyl (JMPR 2002)	serum alkaline phosphatase activity serum AST		260 100	www.inchem.org/documents/jmpr/jmpmono/2002pr09.htm
Cyprodinil (JMPR 2003)	spongiosis hepatitis	5.1		www.inchem.org/documents/jmpr/jmpmono/v2003pr03.htm
Famoxadone (JMPR 2003)	cataracts microscopic lenticular degeneration	29 29		www.inchem.org/documents/jmpr/jmpmono/v2003pr05.htm
Tributyltin (EFSA 2004)	testis weight		9.1	www.efsa.europa.eu/EFSA/efsa_locale-1178620753812_1178620762916.htm
Fumonisin (EFSA 2005)	nephrosis	8.6		www.efsa.europa.eu/EFSA/efsa_locale-1178620753812_1178620807204.htm
Deoxynivalenol (EFSA 2004)	body weight		10.5	www.efsa.europa.eu/EFSA/efsa_locale-1178620753812_1178620763160.htm
Ethyl lauroyl arginate (EFSA 2007)	white blood cell counts		23	www.efsa.europa.eu/EFSA/efsa_locale-1178620753812_1178622334379.htm

(a): As calculated by the Scientific Committee.
(b): Two-sided 90%-confidence interval for extra risk was calculated by the likelihood profile method
(c): Two-sided 90%-confidence interval was calculated for the difference on log-scale, and then transformed back, resulting in the confidence interval for percent change (see Slob 2002 for further statistical assumptions).

The BMD approach involves a statistical method, which uses the information in the complete dataset instead of making pair-wise comparisons using subsets of the data. In addition, the BMD approach can interpolate between applied doses, while the NOAEL approach is restricted to these doses. Therefore, a BMDL is always associated with a predefined effect size for which the corresponding dose has been calculated, while a NOAEL represents a predefined dose and the corresponding potential effect size is mostly not calculated. Therefore, a BMDL value gives more information than a NOAEL, by explicitly indicating the upper bound of effect at that dose as defined by the BMR.

An inherent consequence of the BMD approach is the evaluation of the uncertainty in the calculated BMD, which is reflected by the confidence interval around the BMD. This is a difference with the NOAEL approach that is particularly important in cases where the data are limited. Although in such cases a NOAEL can often be derived, the uncertainty in the value obtained remains unknown.

The data requirements of the NOAEL approach for the purpose of risk assessment have been incorporated into internationally agreed guidelines for study design, e.g. OECD guidelines for the testing of chemicals. However, the utility of the data depends not only on these global aspects regarding study design (e.g. number of dose groups, group sizes), but also on aspects of the quality of the specific study, such as actual doses selected and variability in the responses observed. While in the NOAEL approach, the utility of the data is based to a considerable extent on a priori considerations such as study design, a BMD analysis is less constrained by these factors, as discussed above. In addition, it goes further, by evaluating the data taking the specifics of the particular dataset into account (e.g. the scatter in the data, dose-response information). In this way, a more informed decision on whether a dataset is acceptable for deriving the RP is possible. It should be noted that the BMD confidence interval has already accounted for the limitations of the particular dataset, so that data limitations (e.g. sample size) is a less crucial issue than it is for the NOAEL.

Although the current international guidelines for study design have been developed with the NOAEL approach in mind, they offer no obstacle to the application of the BMD approach. The current guidelines may, however, not be optimal given that the BMD approach allows for more freedom in balancing between number of dose groups and group sizes (Slob 2014). As these guidelines are revised, e.g. within the OECD Test Guidelines Programme, the possibility to recommend study designs that tend to result in better dose-response information (e.g. more dose levels with the same total number of animals) should be taken into account.

2.3.4. NOAEL and BMD approach: some illustrations

This section provides some illustrations of the NOAEL and BMD approaches to dose-response assessment. In the first and second example, real dose-response data from toxicity studies are used to illustrate the NOAEL approach vs. the BMD approach, in the case of continuous and quantal response data, respectively. The third example relates to human (observational) dose-response data.

Example 1: Continuous dose-response data

This example relates to body weights measured in a subchronic NTP study. The BMR in continuous responses should be interpreted as a measure of the degree or severity of the effect, as opposed to the BMR in quantal data which reflects a change in incidence (see example 2).

To illustrate the differences between NOAEL and BMD approaches, both will be applied to this particular dataset. In the NOAEL approach each dose group is compared with the response in the control group, and, as shown in the last column of Table 2, effects at doses of 215 and 419 mg/kg are statistically significantly different at $p < 0.05$, while the other doses are not. Based on the criterion of a statistically significant result, 76 mg/kg would be designated as the NOAEL. Nonetheless, the upper 95%-confidence bound (one sided) of the effect that could occur at this dose level is a 4.7% decrease in body weight (Table 2).

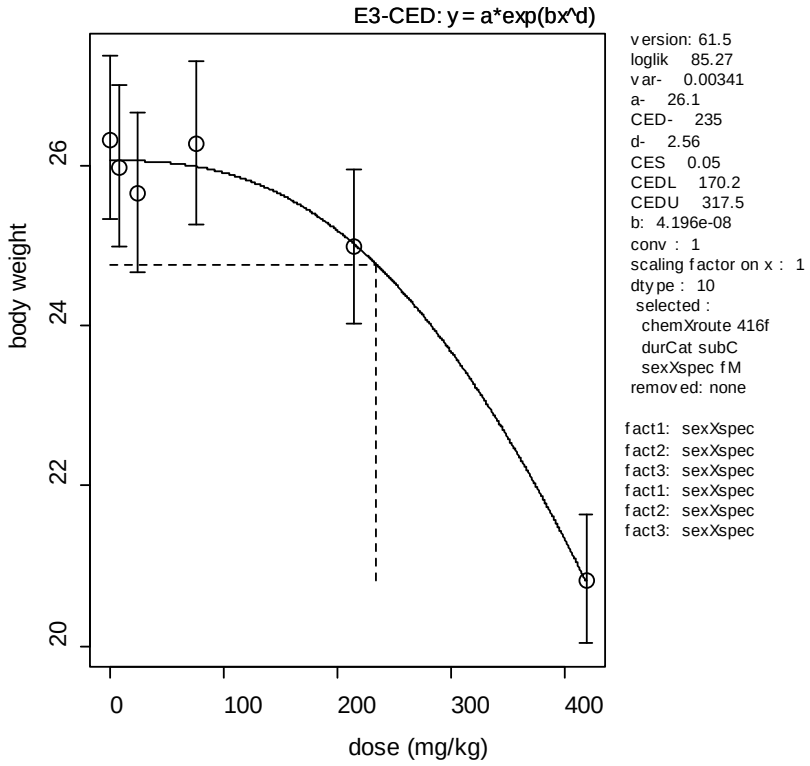
377 **Table 2:** Pair-wise comparison of dose groups, data from Fig. 2

Dose (mg/kg bw)	N	Geometric mean (g)	ES (%)	Lower 95%-confidence bound (one sided) of ES (%)**	Upper 95%-confidence bound (one sided) of ES (%)**	t-statistic	p-value
0	10	26.3					
8	10	26.0	- 1.3	- 5.7	3.3	0.491	0.31
25	10	25.7	- 2.6	-6.9	2.0	0.962	0.17
76	10	26.3	- 0.24	-4.7	4.4	0.087	0.47
215	10	25.0	- 5.1	-9.4	-0.71	1.93	0.029
419	10	20.8	- 21	-25	-17	8.64	0.000

ES = effect size (in percent change compared to response at dose zero)
** Two-sided 90%-confidence interval was calculated for the difference on log-scale, and then transformed back, resulting in the confidence interval for percent change (see Slob 2002 for further statistical assumptions)

378 To illustrate the BMD approach for the same dataset, a dose-response model ($y = a \exp(b\, x^d)$) was
379 fitted to the data, and a BMR representing a 5% decrease in body weight was used (see Fig. 2). The
380 output of this model results in a $BMDL_{05}$ (at BMR = 5%) of 170 mg/kg (see legend of Fig. 2).

381 In this dataset the $BMDL_{05}$ is higher than the NOAEL (170 vs. 76 mg/kg). Nonetheless, it can be
382 stated that the effect size at the $BMDL_{05}$ of 170 mg/kg is smaller than 5% (with 95% confidence).
383 Note that the pair-wise comparison (see Table 2) led to the conclusion that the effect size at 76
384 mg/kg is smaller than 4.7% (again with 95% confidence), similar to the BMR used for the BMDL of
385 170 mg/kg. For the BMD approach to result in a BMDL similar to the NOAEL of 76 mg/kg, the BMR
386 needs to be set at 1.3% in this dataset. In other words, while the NOAEL can only state that effects
387 smaller than 4.7% are unlikely, the BMD approach can state that effects smaller than 1.3% are
388 unlikely, at the same dose, and using the same data. This greater precision illustrates that the BMD
389 approach makes better use of the information in the data by analysing the complete dataset, rather
390 than making comparisons between single dose groups and the control group.



391

392 **Figure 2:** Body weights in ten individual animals per dose plotted against dose in mg/kg bw (data from NTP
393 study 416).

394 Circles represent (geometric) group means, with 90%-confidence intervals. The solid curve is the fitted dose-response model
395 using PROAST v. 61.5. The dashed lines indicate the BMD at a BMR of 5%. CED = BMD, CEDL = BMDL, CEDU = BMDU.
396

397 **Example 2: Quantal response data**

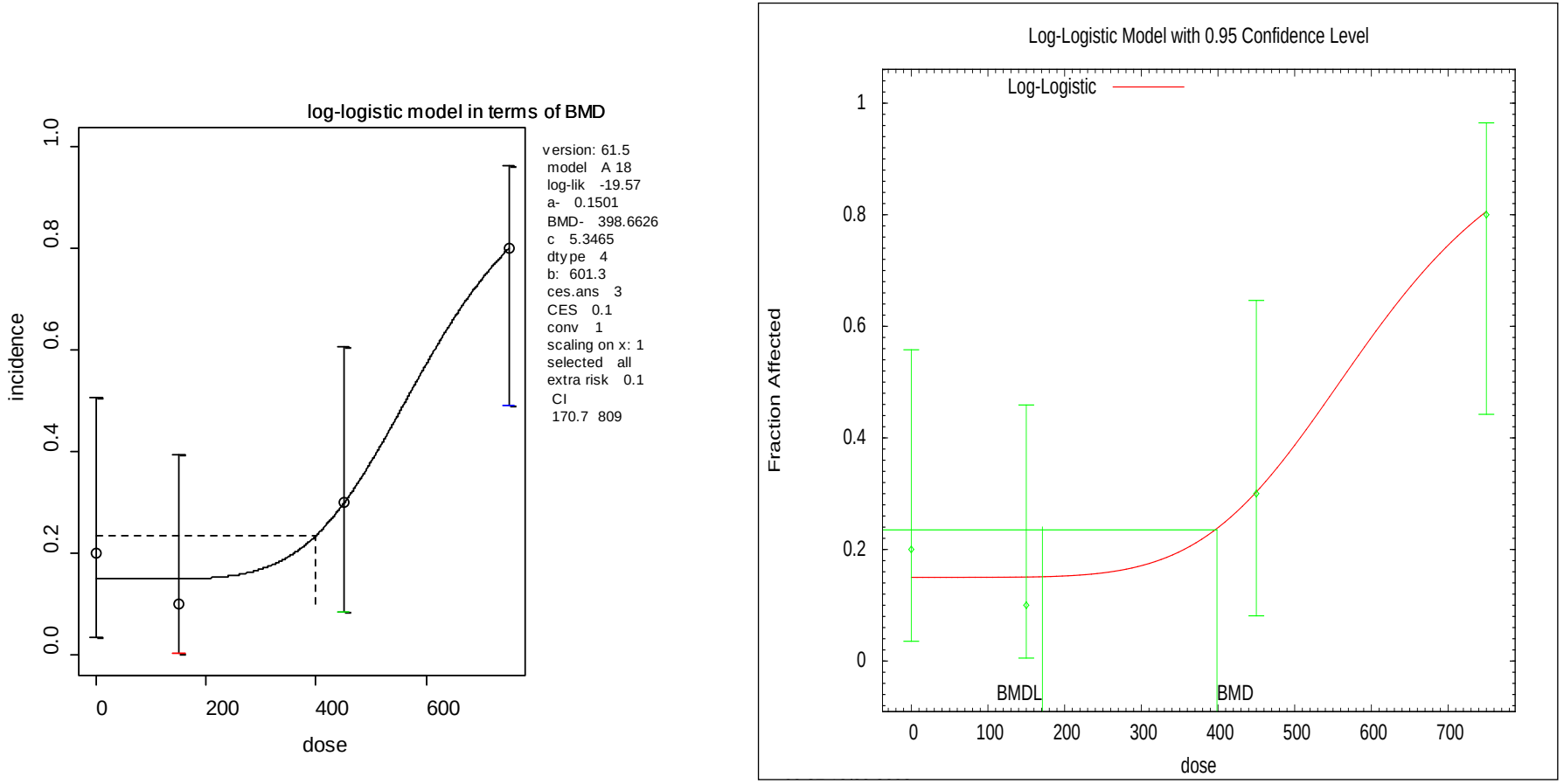
398 For quantal data, the BMR is defined as a specified increase in incidence over background. A BMR of
399 10% (extra risk) is used in the following example, illustrated in Figure 3.

400 Here, the mid-dose and low-dose incidences are not statistically significantly different from the
401 background response. Hence, the middle dose of 450 mg/kg is the NOAEL for this endpoint in this
402 study. In this case, a pair-wise comparison with the background response results in a very large upper
403 95%-confidence bound (one sided) for the effect size at the NOAEL: an extra risk[7] value of around
404 47%.

405 Modelling the dose-response data (see Fig. 3) using a log-logistic model as an illustration results in a
406 BMD_{10} of 399, and a $BMDL_{10}$ of 171 mg/kg, considerably lower than the NOAEL.

407 The BMD approach allows for the statement that the associated effect at the BMDL is not greater than
408 10% (with 95% confidence), which is considerably lower than the upper bound of effect of around
409 47% at the NOAEL, as calculated based on a pair-wise comparison of the background response and
410 the NOAEL dose group.

[7] The upper 95%-confidence bound (one sided) for extra risk was estimated by the likelihood profile method, using the data in the controls and at the NOAEL only, i.e., without using an assumed dose-response model.



411

412 **Figure 3:** Analysis of quantal data as obtained by PROAST and BMDS software.

413 Fraction of affected animals in a toxicity study with 10 animals in each dose group. A dose-response model has been fitted to the data (solid curve) and the horizontal line indicates the BMR of 10%
414 extra risk compared to the response at zero dose (according to the curve). Log-logistic model was fitted by PROAST (version 61.5) and BMDS (version 2.6) (see Table 3); the figures presented
415 reflect the way in which the software generates the graphs.

Example 3: Human dose-response data.

The analysis of human dose-response data is generally more complicated than that of typical dose-response data from animal studies, due to confounders that need to be accounted for, and imprecision in the exposure estimates. The example provided here does not deal with these complexities and aims only to illustrate one particular aspect of human data that may occur, that of very small exposure groups. In specific cases, exposure levels are estimated for each individual person. The NOAEL approach could then only be applied if the doses are lumped into a limited number of dose categories. However, such would result in a loss of information. In contrast, the BMD approach can be applied without categorisation, as illustrated in Fig. 4. In this example, every person was scored as showing either normal (=0) or abnormal (=1) eye-hand coordination. It is hard to detect any dose-response relationship by visual inspection for these types of observations. It is however feasible to fit a dose-response model to these data, and demonstrate the existence of a dose-related response. In this example, the curve associated with the fitted model represents the probability of any person responding at a given exposure level. The fitted model resulted in a statistically significant improvement of the fit compared with a fitted horizontal line, indicating that there is a statistically significant effect of the exposure. The BMD approach uses this curve to estimate the exposure level where the extra risk is 10% (see section 2.5.7), together with the BMD confidence interval.

The example further illustrates that the BMD approach may apply in situations without any controls: the background response level can in principle be estimated by the fitted dose-response curve, while the confidence interval for the estimated background indicates how well it could be estimated, given the data available.

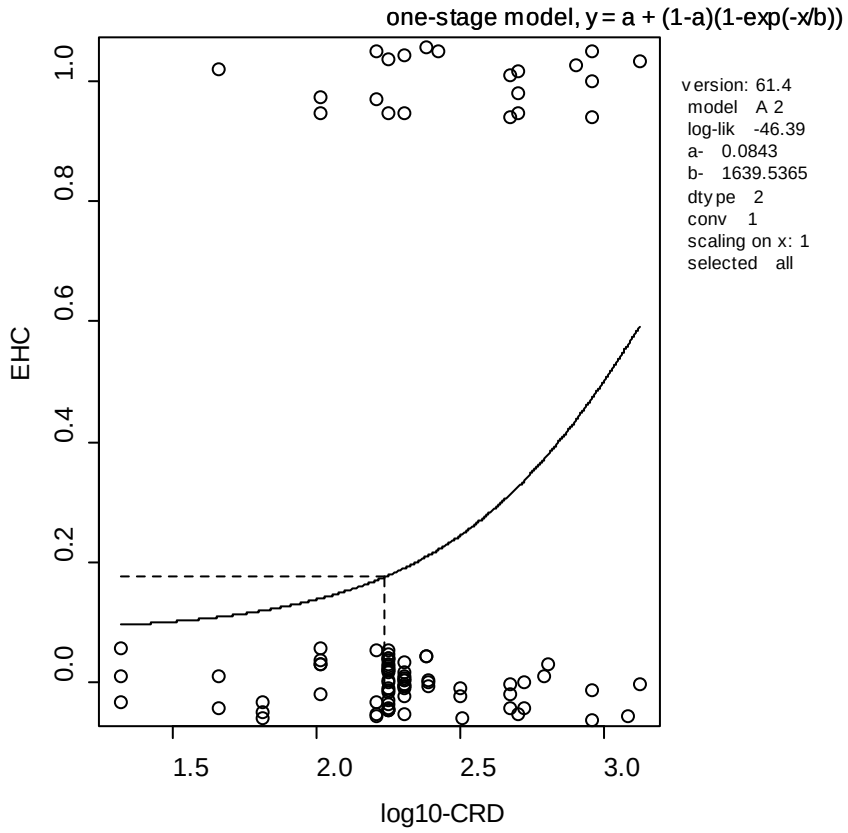


Figure 4: BMD analysis of human dose-response data with individual exposures.

Observed eye-hand coordination scores (0.0 = normal, 1.0 = abnormal) in individual workers (plotted as circles with some artificial vertical scatter to make the ties visible for individuals having the same exposure) as a function of exposure (CRD). A dose-response model has been fitted to these data using PROAST v. 61.4; the BMD_{10} (see dashed lines) was 173, and the $BMDL_{10}$ was 92. A BMR of 10% extra risk was used.

2.4. Consequences for hazard / risk characterisation

In the previous section the BMD approach has been introduced in the context of deriving a reference point (RP). This reference point will be used in hazard characterisation for establishing health-based guidance values, such as acceptable daily intakes (ADIs) for food additives and pesticide residues, and tolerable daily intakes (TDIs) or tolerable weekly intakes (TWIs) for contaminants. It will also be used in risk characterisation of substances that are both genotoxic and carcinogenic, i.e. in establishing margins of exposure (MOEs).

2.4.1. Establishing health-based guidance values

In establishing a health-based guidance value (HBGV) such as an ADI or TDI from a RP, uncertainty factors are applied to the NOAEL (WHO 1987). It has been suggested that larger or additional uncertainty factors might be appropriate when a BMDL is used as the RP. The argument used is that the BMDL does not reflect a "no-effect" dose, in contrast to the NOAEL. This argument is based on the false assumption that a NOAEL is associated with the complete absence of any adverse effect. As discussed (in Section 2.5.2), the default values of the BMR are such that the BMDL on average coincides with the NOAEL. Further, it was shown in section 2.3.3 that the potential magnitude of the effect at the NOAEL can be even greater than the specified effect size (BMR) associated with the BMDL. Taking these considerations into account, an additional uncertainty factor, beyond those normally applied is not necessary (it might actually be argued that an additional uncertainty factor would be needed when using the NOAEL rather than the BMDL, see IPCS 2014). The health-based guidance value derived from the BMDL can be expected to be as protective as the one derived from the NOAEL, i.e., on average over a large number of risk assessments. In conclusion, the default values for uncertainty factors currently applied to the NOAEL are equally applicable to the BMDL.

In some studies, there may be an effect at the lowest dose tested which is statistically significantly different from the response in the control group and biologically relevant (LOAEL). In the NOAEL approach, the LOAEL is traditionally divided by an additional uncertainty factor. However in the BMD approach, it is usually possible to derive a BMDL from such data at the desired BMR and there would be no need for such an additional uncertainty factor. If the desired BMR would imply substantial extrapolation outside the observed dose-range from the fitted model (see section 2.5.2), then a higher BMR can be selected but an additional uncertainty factor to the BMDL may be necessary.

2.4.2. Risk assessment of substances which are both genotoxic and carcinogenic

The Scientific Committee (EFSA, 2005) concluded that, from the options considered, the MOE approach would be the most appropriate one in the risk assessment of substances that are both genotoxic and carcinogenic. They proposed to use the $BDML_{10}$ as the reference point (RP), i.e. the $BMDL_{10}$ should constitute the numerator of the MOE.

2.4.3. Potency comparisons

Comparisons of the potencies of different substances, or of the same substance under different exposure conditions, require information on the doses necessary to produce the same size of effect/response. The BMD approach is a suitable tool for such analyses, as it enables the estimation of equipotent doses by interpolation between applied doses. For the same reason the BMD approach is

also suitable for the derivation of Relative Potency Factors (RPFs) or Toxic Equivalency Factors (TEF) for individual substances in a mixture that share a common mode of toxicological action. The BMD approach has been used to provide relative potency estimates for different organophosphates (Bosgra et al., 2009). Relative potency estimates obtained using the BMD approach are also more appropriate than NOAELs for use in mode of action analyses (Boobis et al., 2006). More recently, the BMD approach has been used for estimating equipotent doses in *in vivo* and *in vitro* genotoxiticy tests which so far has only been used for (qualitative) hazard identification (e.g. Soeteman-Hernandez et al. 2015a and 2015b, Bemis et al. 2015, Wills et al. 2015). Further, the BMD approach can be used for testing if dose addition applies in chemical mixtures (Kienhuis et al. 2015).

2.4.4. Probabilistic risk assessment

Probabilistic approaches in risk assessment are receiving increasing attention, regarding both exposure assessment (e.g. Gibney and Van der Voet, 2003; Fryer et al., 2006; Tressou et al. 2004) and hazard characterisation (e.g. Baird et al. 1996; Swartout et al. 1998; Van der Voet and Slob 2007; IPCS 2014; Chiu and Slob 2015). The BMD approach is compatible with probabilistic hazard characterisation, as the uncertainty in the BMD can be quantified in the form of a distribution (Slob and Pieters, 1998). Further, the dose-response modelling behind the BMD approach provides a means of estimating the magnitude of a potential health effect in the human population, given a particular exposure level (e.g. the current exposure in the population). This has been done, for example, for the mycotoxin deoxynivalenol (Pieters et al., 2004), and for a number of genotoxic carcinogens (Slob et al. 2014).

2.4.5. BMDL vs. NOAEL: Perception of safety

It has been argued that the introduction of the BMD approach may raise problems in communication with risk managers, politicians, consumer organisations and the public because the BMDL is perceived as an effect level. On the other hand, the NOAEL is sometimes perceived incorrectly as a level that is without any effects. However, as explained in section 2.4.1, use of the BMDL in risk assessment does not fundamentally change the basic approach or assumptions.

An argument in favour of the BMD approach is that this approach provides a higher level of confidence in the conclusions in any individual case since the BMDL takes into account the statistical limitations of the data better than the NOAEL. This does not imply that re-evaluation of all previous data is needed, because as stated in section 2.4.1, the NOAEL and BMDL are expected to be similar on average. A re-evaluation would certainly not be necessary in circumstances where large margins exist between the estimated daily intake and the health based guidance value, e.g. ADI. For substances where the actual estimated daily intake appears to be close to or exceeding the health-based guidance value, a refined risk assessment might result from a re-evaluation of the data, using the BMD approach.

It also has to be recognised that there are a number of sources of uncertainty in a risk assessment, and dose-response modelling is only one of these. In assessing the likely benefits of applying the BMD approach in a given risk assessment, some consideration should be given to the sources of uncertainty, their magnitude and the likely impact in the assessment. The latter can be done in a probabilistic way, as recommended by IPCS (2014). Such information will help to determine whether the likely refinement provided by the BMD approach will result in a substantial change in the risk assessment.

In addition, when a health-based guidance value is based on the BMD approach, it takes into account all the data from the dose-response curve. The BMD method provides a better basis to quantify the risk in situations where the health-based guidance value is exceeded, and, thus, is a better basis for risk communication.

Finally it is important to realize that health-based guidance values like the ADI or TDI are defined as a level to which an individual may be exposed daily over his or her lifetime without appreciable health risk, and this definition does not change when the health-based guidance value is derived from a BMDL instead of a NOAEL.

2.5. Guidance to apply the BMD approach

This section provides an overview of how to derive a BMD confidence interval from dose-response data and recommendations are given on particular choices to be made. The guidance refers not only to *in vivo* data but could be applied also to other types of data (e.g. *in vitro* data). Although currently available software allows for the application of the BMD approach without detailed knowledge of computational technicalities, a conceptual understanding of the method, as described in this opinion, is a prerequisite for correct interpretation of the results.

The application of the BMD approach may be summarized as a process involving the following steps:

1. Specification of type of dose-response data (section 2.5.1)
2. Specification of the BMR (section 2.5.2)
3. Selection of candidate dose-response model(s) (section 2.5.3 and 2.5.4)
4. Fitting the candidate models and calculate the BMD confidence interval for each model (section 2.5.5)
5. Combining the results from the various models into one single BMD confidence interval, with the lower bound (BMDL) as the RP (section 2.5.7).

These steps are further discussed below.

2.5.1. Specification of type of dose-response data

Response data may be of various types, including continuous, quantal and ordinal. The distinction between data types is important for statistical reasons (such as assumption of underlying statistical distribution), but also for the interpretation of the BMR. See section 2.3.4 (examples 1 and 2) for the interpretation of the BMR in continuous and in quantal data. Ordinal data may be regarded as an intermediate data type: they arise when a severity category (minimal, mild, moderate, etc.) is assigned to each individual (as in histopathological observations). Ordinal data could be reduced to quantal data, but this implies loss of information, and is not recommended.

For continuous data, the individual observations should ideally serve as the input for a BMD analysis. When no individual but only summary data are available, the BMD analysis may be based on the combination of the mean, the standard deviation (or standard error of the mean), and the sample size for each treatment group. Using summary data may lead to slightly different results compared with using individual data (Slob 2002; Shao et al. 2013). For quantal data the number of affected individuals and the sample size are needed for each dose group.

2.5.2. Specification of BMR

The BMR is a specific value of the effect size selected for estimating the associated dose (the “true” BMD). The meaning of an effect size is different in continuous vs. quantal endpoints, and are therefore discussed separately.

Continuous data

For continuous data the metric for the BMR could be defined in various ways. One option is to define the BMR as a change in the mean response relative to the variation in the control group, as measured

by the standard deviation (SD). The US EPA Benchmark Dose Technical Guidance (2012) recommends to always report the estimated BMD associated with BMR in terms of a change in means equal to 1 SD. However, one of the weaknesses of this definition of BMR is that the associated (true) BMD then depends on the particular study, due to study-specific factors (measurement error; dosing error; heterogeneity in experimental conditions). Another problem of using the 1 SD metric is that the associated BMD estimate cannot be translated into an equipotent dose in populations with larger within-group variation, including humans. Another option is to define the BMR as a percent change in mean response; the BMD associated with such a BMR does not depend on the within-group variation and therefore is more stable among different studies examining the same dose-response, as well as among different populations. Therefore, the Scientific Committee recommends to define the BMR as a percent change in the mean response as compared to the background response.

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 (see Fig. 5), while in most individual datasets they differed within one order of magnitude. Similar observations have also been made in studies of foetal weight data (Kavlock et al., 1995). While the BMR of 5% for continuous data is recommended as a default, it might be modified based on toxicological or statistical considerations. For example, a 20% change in a liver enzyme in serum might still be considered sufficiently small for deriving an RP, based on biological considerations. As a statistical consideration, one might consider to select a BMR higher than 5% for endpoints that tend to show a relatively large within-group variation (in terms of coefficient of variation), and/or a relatively high maximum response (if known, based on experience with that endpoint over a larger number of studies (Slob and Setzer, 2014)). Increasing the BMR (in terms of a percent change) for data showing a relatively large maximum response is somewhat similar to using a BMR defined as a change equal to 1 SD (the default metric for the BMR in the BMDS software); the main difference is that the BMR expressed in terms of a percent change allows for comparison among studies and populations that differ in within-group variation.

Quantal data

For quantal data the BMR is defined in terms of an increase in the incidence of the lesion / response scored, compared with the background incidence. The common way of doing this is either by additional risk (incidence at a given dose minus incidence in the controls), or extra risk, i.e. the additional risk divided by the non-affected fraction of the population (see section 2.3.3, footnote 5). Epidemiologists more often use relative risk, where a given incidence (prevalence) is divided by the control incidence.

For quantal response data observed in experimental animals BMR values of 1%, 5% or 10% (extra or additional risk) were initially proposed (Crump, 1984; EPA, 1995). Various studies estimated that the median of the upper bounds of extra risk at the NOAEL was close to 10%, suggesting that the $BMDL_{10}$ may be an appropriate default (Fowles et al., 1999, Allen at al., 1994, Sand et al., 2011). Also, a BMR of 10% appears preferable for quantal data because the BMDL can become substantially dependent on the choice of dose-response model at lower BMRs (Sand et al., 2002).

In conclusion, for experimental animal studies, the Scientific Committee proposes that a default BMR value of 5% (change in mean response) be used continuous data and 10% (extra risk) for quantal data. As stated previously, the default BMR may be modified based on statistical or biological considerations. For example, if the BMR is considerably smaller than the observed response(s) at the lowest dose(s), leading to the need to extrapolate substantially outside the observation range, a larger BMR may be chosen. The biological relevance of the new BMR value should be discussed. An additional assessment factor may be needed to account for the higher BMR, depending on the biological impact of the latter. The rationale for deviating from the default BMR should be described and documented.

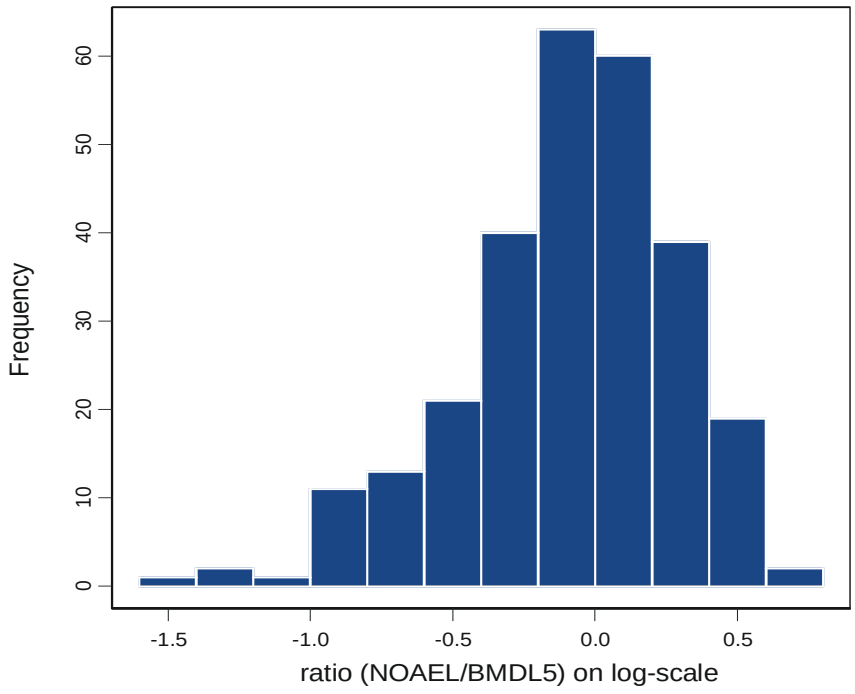


Figure 5: Histogram of 395 NOAEL/BMDL05 ratios (log10-scale) for the same dose-response data in rat and mouse (NTP) studies (Bokkers and Slob, 2007).

The BMDL05 relates to a BMR of 5%. Six endpoints were considered: BW, relative and absolute liver and kidney weight, red blood cell counts. The geometric mean of the ratios is close to 1, i.e. on average the NOAEL is similar to the $BMDL_{05}$.

2.5.3. Recommended dose-response models

In the current opinion, the term dose-response model is used for a mathematical expression (function) that describes the relationship between (mean) response and dose. This section will deal with dose-response models in that sense. The distributional part of dose-response models will be discussed in section 2.5.4.

Before discussing the dose-response models that may be used in a BMD analysis, it is important to understand that the purpose of a BMD analysis is not to find the best estimate of the (true) BMD but rather to find all plausible values of the (true) BMD, given the data available. This question will be answered as follows. A dose-response model can produce different curves by varying the parameters of that model. Based on a statistical criterion, the curves that are compatible with the data are selected. For each of those curves the dose associated with the BMR is calculated, and together these doses comprise a confidence interval for the (true) BMD, with a specified confidence level (recommended default: 90%). This procedure is repeated for a number of models, since different models might produce other curves that are compatible with the data as well (this is called "model uncertainty"). Each model considered will result in a BMD confidence interval, and by combining those intervals the result will be a BMD confidence interval that takes "model uncertainty" into account. Thus, finding the single "best" fitting model for a given dataset is not the aim of a BMD analysis; the aim is to evaluate the uncertainty in the BMD resulting from that dataset, including model uncertainty.

For this reason, the appropriate approach is to fit various models to the dataset considered. This section provides a discussion of candidate models, in particular those that are currently available in the BMD software packages (e.g. BMDS, PROAST). These models are suitable for analysing toxicological datasets in general. If other software is used, it is recommended to apply the same set of candidate models.

Table 3 summarizes the recommended models. As can be seen from the table, the models for continuous or quantal data differ; they will be discussed below. There are, however, two special models that relate to both types of data: the so-called full (or saturated) model and the null model. The full model describes the dose-response relationship simply by the observed (mean) responses at the tested doses, without assuming any specific dose-response. It does however include the (same) distributional part of the model (see next section) and thus it may be used for evaluating the goodness of fit of any dose-response model (see section 2.5.5). The null model expresses the situation that there is no dose-related trend, i.e. it is a horizontal line, and may be used for statistically evaluating the presence of a dose-related trend (see section 2.5.7).

Models for continuous data

For continuous data both the exponential family and the Hill family of models are recommended. These models have the following properties:

- they always predict positive values, e.g. organ weight cannot be ≤ 0
- they are monotonic (i.e. either increasing or decreasing),
- they are suitable for data that level off to a maximum response,
- they have been shown to describe dose-response datasets for a wide variety of endpoints adequately, as established in a review of historical data (Slob and Setzer, 2014),
- they allow for incorporating covariates in a toxicologically meaningful way (see section 2.5.5).
- they contain up to four parameters, which have the same interpretation in both model families, in particular: *a* is the response at dose 0, *b* is a parameter reflecting the potency of the chemical (or the sensitivity of the population), *c* is the maximum fold change in response compared to background response, and *d* is a parameter reflecting the steepness of the curve (on log-dose scale). The four parameters are summarized in Fig. 6.

The Scientific Committee recommends more parametric dose-response models with the above characteristics to be developed for continuous data.

For both the exponential and the Hill family of models, Table 3 presents for each family two different models, respectively: one with three parameters and one with four parameters. The previous guidance (EFSA, 2009) included for each family two other members, but these are no longer recommended, as BMD confidence intervals tend to have low coverage[8] when parameter *d* is in reality unequal to one.

[8] A confidence interval has low coverage when it does not include the true value of the parameter (e.g. BMD) with the probability that is implied by the confidence level. For example, a two-sided 90% confidence level should miss the true value with probability 10%.

692 **Table 3:** Expressions of the recommended models for use in the BMD approach, with (mean)
693 response (y) being a function of dose (x), both on the original scale. See Table 5 in
694 Appendix A for the equivalent model expressions used in BMDS software.

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	
Null Model[ii)]	1	y = a	a>0 for continuous data 0<a<1 for quantal data
Continuous data			
Exponential family			
3-parameter model[9]	3	$y = a \exp(bx^d)$	a>0, d>1
4-parameter model[10]	4	$y = a [c – (c-1)\exp(- bx^d)]$	a>0, b>0, c>0, d>0
Hill family			
3-parameter model[9]	3	$y = a [1 - x^d/(b^d+x^d)]$	a>0, d>1
4-parameter model[10]	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>0
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>0
Gamma	3	$y = a + (1-a) \text{CumGam}(bx^c)$	$0 \leq a \leq 1$, b>0, c>0
LMS (two-stage) model	3	$y = a + (1-a)(1- \exp(– bx – cx^2))$	a>0, b>0, c>0

695 a, b, c, d: unknown parameters that are estimated by fitting the model to the data.
696 CumNorm: cumulative (standard) normal distribution function.
697 CumGam: cumulative Gamma distribution function.
698 [i)] the full model will result in the maximum possible value of the log-likelihood (given the statistical assumptions) for the dataset
699 considered.
700 [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-
701 response (= horizontal line).

[9] Called model 3 in the BMDS and PROAST software
[10] Called model 5 in the BMDS and PROAST software

In the model expressions for continuous data, parameter *a* (reflecting the background response) is included multiplicatively, in line with defining the BMR as a percent change (rather than a difference) compared to background response. Further, it matches the common way of normalizing responses in different subgroups to 100% response. Occasionally, dose-response data may be expressed such that they include negative values, for instance body weight gains decreasing from positive to negative values at high doses. In those cases, the recommended models that are strictly positive are no longer valid and models with an additive background parameter would be needed. Preferably, however, the body weight gains should be expressed as ratios (percent changes) rather than differences, if the individual body weight data are available.

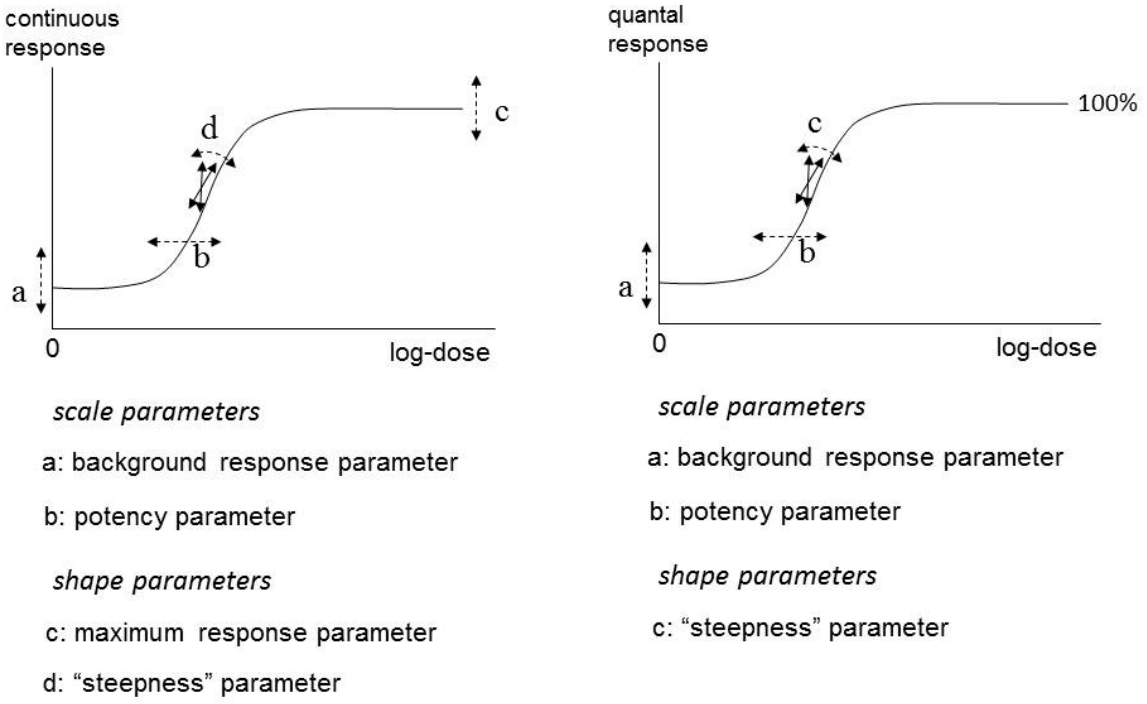


Figure 6: The four model parameters a, b, c, and d and their interpretation for continuous and quantal data. The dashed arrows indicate how the curve would change when changing the respective parameter.

The US EPA BMDS includes some additional models for continuous data, in particular, the power model and the polynomial (including the linear) model. These models are additive with respect to the background response, which could result in fitted curves predicting negative values. Therefore, the Scientific Committee does not recommend using these models.

Quantal data

Table 3 lists seven models that are recommended to be used for quantal data. The two-stage model is a member of the nested family of linearized multi-stage models (LMS). The two-stage model is recommended to be used from this family as it has next to the scale parameters (*a* and *b*) one single shape parameter (c), just like most other quantal models. Furthermore, general experience has shown that the three-stage model (recommended in the previous version of this guidance document) rarely provides a better fit to the data; consequently, this model has now been removed from the table of recommended models.

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.

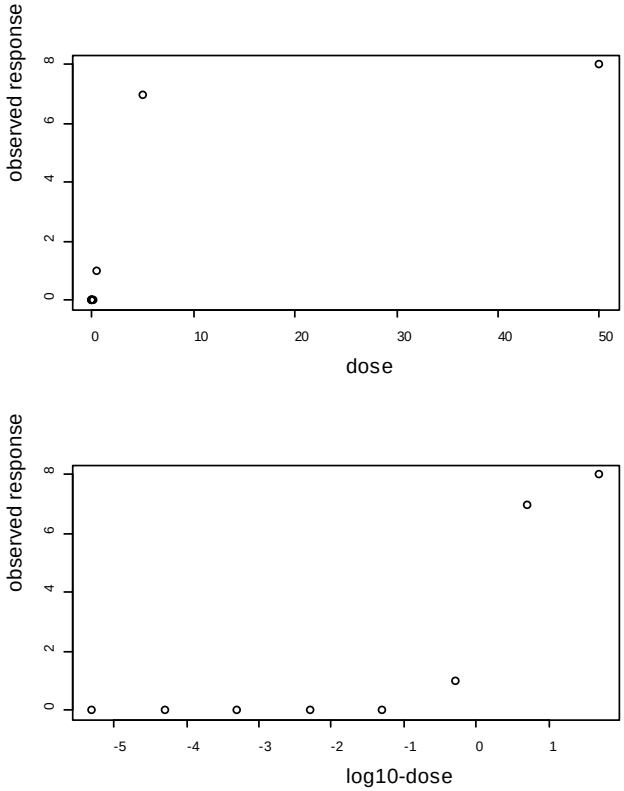


Figure 7: A dose-response dataset where the response is plotted against the dose (upper panel) and against the log-dose (lower panel).

The slope appears infinite when the response is plotted against the dose, while it appears to be "threshold-like" when plotted against the log-dose. The lower doses are squeezed to dose zero when plotted against dose, and hence not visible. When plotted against log-dose they become visible, showing that in reality there is a large range of doses with virutally no effect.

2.5.4. The distributional part of dose-response models

The dose-response models introduced and discussed in the previous section describe the mean response as a function of dose, assuming that there was no random sampling error. As this is unrealistic, the dose-response model also needs to describe the within-(dose-)group variation. This may be called the "distributional part" of the dose-response model.

In continuous data, the within-group variation can be directly observed as the scatter of the individual data around the fitted curve at the respective dose. This scatter can be characterized by a statistical distribution. Basic choices include two-parameter distributions such as the log-normal, gamma, Weibull, inverse-Gaussian distribution etc. Although different choices can be considered for different datasets, the log-normal distribution has shown to be an appropriate choice across a variety of datasets, and therefore is proposed as the default choice. When individual data are available, the Scientific Committee recommends to check the log-normality assumption (e.g. by using a normal probability plot - also called QQ-plot) and to consider alternative distributions, if considered necessary. However, one should always realize that a deviation from lognormality could also be the result of another misspecification of the distributional part of the model (e.g. litter effects were not taken into account), or by specific problems in the data (e.g. outliers).

The lognormal distribution has two parameters: the mean and standard deviation of the response on the log-scale at a given dose. Taking the exponential function (depending on the base of the log used) of the mean and the standard deviation results in the geometric mean (= median) and geometric standard deviation on the original response scale, respectively (Slob, 1994). With the

default lognormality assumption, the models in Section 2.5.3 describe the geometric mean as a function of dose. In the default application of these models, it is assumed that the standard deviation for the log-responses (or, equivalently, the geometric standard deviation) does not depend on dose. When there is evidence that the geometric standard deviation does change with dose, an extended model, which considers dependence of the standard deviation on dose can be used. Ignoring that dependency (while in reality it exists), the fitted dose-response model for the mean and the BMD estimate are in general expected to be still appropriate. However, the standard errors of the parameter estimates are expected to be larger, resulting in lower and hence more conservative BMDLs. For this reason, and because it is not easy to decide whether the within-group variation does indeed depend on dose (given the impact of sampling error on the observed standard deviations) the assumption of constant variability among dose groups is recommended as the default practice.

In quantal data, the within-group variation is normally not directly visible, given that most quantal datasets only have one observed incidence per dose. Yet, this observed incidence is subject to random sampling error just as well. When the experimental units do not show any dependencies (like litter effects) the sampling error in observed incidences will be binomially distributed. Therefore, the binomial distribution is the default assumption for quantal data.

When there are litter effects, they need to be taken into account. Ignoring them will result in too small BMD confidence intervals. One way to take litter effects into account is by assuming an additional (e.g. Beta) distribution that describes the variation among the dams. In PROAST all usual quantal models can be fitted while taking the litter effects into account based on that principle. In BMDS three quantal models are available that can account for litter effects.

2.5.5. Fitting models

The currently available BMD software from US EPA and RIVM takes care of fitting a model, which means finding the values of the unknown parameters in the model that make the associated dose-response curve approach the data as closely as possible. This is called the best fit of that model and is achieved by maximizing the log-likelihood that can be reached by that model.

Convergence

The available BMD software fit the recommended models by applying numerical algorithms, which are optimization procedures: the fit of the model is re-evaluated over and over again for different values of the parameters, until the log-likelihood can no longer be improved. If the algorithm is able to find the maximum likelihood, while this holds for just one set of parameter values, the software will report that the algorithm has “converged”. It may happen, however, that the algorithm reports that no convergence was reached. There could be various reasons for that, but usually this indicates that the data did not provide sufficient information to appropriately estimate all the parameters in the model. For instance, there may be different sets of parameter values that would result in similar log-likelihood values. Clearly, this would hamper the establishment of the statistically best estimate of the BMD, but for risk assessment purposes the BMD confidence interval is of interest. Simulations showed that convergence may not be critical in providing a reliable BMD confidence interval, and therefore a message of non-convergence does not necessarily imply that the model should be rejected. However, non-convergence does typically indicate that the data are not informative enough to estimate all parameters for the model at hand, and this should be considered as an alert.

The AIC criterion

For the purpose of comparing the fit of different models, the AIC (Akaike Information Criterion) is a convenient criterion as it directly integrates the log-likelihood and the number of model parameters in one single value. The AIC is calculated as $-2 \log(L) + 2p$ with $\log(L)$ the log-likelihood of the model,

and p the number of parameters. The first term, -2 log(L) will decrease when the model gets closer to the data. To penalize for the number of parameters, AIC includes the term 2p, which increases the value of AIC when the number of parameters increases. Thus, the model with a relatively low AIC may be considered as providing a good fit without using too many parameters.

According to Burnham and Anderson (2002), different models that result in AICs not differing by more than 2 units may be regarded as describing the data equally well. Further, the full model tends to show the smallest AIC, and the null model the largest, although deviations may occur when there is a large number of dose groups.

The AIC criterion can be used to check if there is statistical evidence of a dose-related trend. For a fitted model to show statistical evidence of a dose-related trend, the Scientific Committee proposes that its AIC should be lower than the $AIC_{NULL}+2$.

The AIC criterion can also be used to compare the fit of any model with that of the full model. Theoretically, the AIC of a fitted model should be no more than 2 units larger than the full model's AIC. If the model with the minimal AIC is more than two units larger than that of the full model ($AIC_{min} > AIC_{FULL} + 2$), this could be due to the use of an inappropriate dose-response model (e.g. it contains an insufficient number of parameters), or to misspecification of the distributional part of the model (e.g. litter effects are ignored), or to non-random errors in the data (see section 2.5.7).

Covariates

Besides fitting dose-response models to single datasets, it is possible to fit a given model to a combination of datasets which differ in a specific aspect, such as sex, species, or exposure duration, but are similar otherwise. In particular, the response parameter (endpoint) needs to be the same. By fitting the dose-response model to the combined dataset, with the specific factor included in the analysis as a so-called covariate, it can be examined in what sense the dose-responses in the subgroups differ from each other, based on statistical principles (like AIC).

In general, there are three possible outcomes of such an analysis. First, it may be found that the subgroups show similar dose-responses, and that a single curve may be used to describe all subgroups combined. Second, the subgroups may be found to differ in dose-response but only partially so. For instance, they may show different background responses (at dose zero) but be equally sensitive to the chemical. Or, they may differ in sensitivity but their dose-responses may otherwise have the same shape. In the latter case, the analysis will result in subgroup-specific BMD confidence intervals. The third possible outcome is where the subgroups appear to differ in all parameters in the model. In this case the result of the combined dose-response analysis will be identical to analysing the subgroups separately. With the appropriate software (e.g. PROAST) a combined analysis can be performed, and will indicate how the combined dataset could be best described.

Combining datasets in a dose-response analysis with covariate(s) may have two reasons. The first is that it provides a powerful method for examining and quantifying potential differences in dose-response between the subgroups. For instance, the problem formulation might indicate that the assessment should specifically focus on sex differences, in which case it would be important to know if the data provide evidence that both sexes actually differ in sensitivity to the test material, and if so, to have a precise estimate of the difference in (true) BMDs between male and female animals. As another example, by combining different chemicals affecting the same endpoint an effective estimate of the relative potencies will be obtained (Bosgra et al, 2009). Or, one might be able to link the more precise information on the potencies of various chemicals to mechanism of action hypotheses (Wills et al. 2015).

The second reason for combining datasets and applying the covariate approach is to improve the precision of the estimated BMD(s), i.e. to obtain a smaller BMD confidence interval. This is particularly relevant when the individual datasets provide relatively poor dose-response information (for an illustration see figure 11 in Slob and Setzer, 2014). As long as at least one of the parameters in the model does not appear to differ among the subgroups, it is useful to include the factor that

discriminates the subgroups as a covariate in the analysis: the common parameter can then be estimated from all data combined, and hence will be known more precisely, resulting in a more precise estimate of the (true) BMD(s).

2.5.6. Model averaging

As discussed in section 2.5.3, the BMD approach does not aim to find the single statistically best estimate of the BMD but rather all plausible values that are compatible with the data. Therefore, the goal is not to find the single best fitting model, but rather to take into account the results from all models applied. The recommended way to do that is by the so-called Model Averaging approach.

Recently, it has been shown that multi-model estimation and inference using model averaging is the best way to account for model uncertainty and at the same time for the uncertainty related to the sampling errors in the data (Burnham and Anderson 2004, Wheeler and Bailer 2007, 2008, 2009). In model averaging, the individual model results are combined by using weights, with higher weights for models that fit the data better. These weights are often defined in terms of the AIC.

Briefly, model averaging consists of two main steps. First, the average response is calculated for a large number of doses, taking the weighted average from the fitted models involved. The BMD can now be calculated for the average model. Second, a large number of artificial datasets are generated based on the average model, and for each dataset the first step is repeated. This results in a large number of BMDs: the lower and higher 5^{th} percentiles define the BMD 90% confidence interval.

In this document, the MADr-BMD program as described in Wheeler and Bailer (2008) has been used to perform example model averaging analyses (see Section 2.5.9, Example 2).

2.5.7. Establishing the BMD confidence interval

BMD confidence interval for a given dataset

The flowchart of Fig.8 shows how to proceed once the models are fitted to the data.

When the software reported “no convergence” for one or more models, this may be taken as an alert: apparently, the data are not very informative, or the model may be over-parameterized. In cases of non-convergence, it is recommended to consult a specialist on dose-response modelling on how to proceed with the BMD analysis. At EFSA level, a standing working group will be established to assist EFSA Experts and Staff on BMD-related issues (see section 4). Preliminary simulations have shown that non-convergence may have little impact on the BMD confidence interval.

The next step consists in checking if at least one of the models revealed a dose-related trend, i.e. one of the model’s AIC is lower than the AIC of the null model + 2 units. If this is not fulfilled, there is no dose-related trend and the BMD analysis can stop.

If nested models have been used (i.e. for continuous data), one single member is selected per model family: the one that resulted in the lowest AIC. Then, from all (non-nested) models, the lowest AIC is determined. If this lowest AIC exceeds the AIC of the full model by more than 2 units, this may be considered as an alert. A first thing to consider in that case is whether the distributional part of the model needs to be adjusted (see section 2.5.4). Then, it should be explored if there might be other problems in the data, e.g. systematic errors caused by some unintended experimental factor differing among dose groups. This would be more likely if the study was not performed according to an appropriate study protocol (e.g. randomizing animals and order of treatments, avoiding cage effects as a confounding factor). Associated non-random data errors may result in an AIC_{min} that is substantially larger than the AIC_{full}. An indication of non-random errors may be found by checking whether the confidence intervals around the (mean) responses at each dose are hit by the fitted dose-response curve; if not, this may indicate that there is more error in the data than expected from random sampling error. If non-random errors appear to be the most likely explanation of the alert, the

930 BMD analysis can in principle continue according to the flowchart, in particular when the impact on
931 the estimated shape of the dose-response is minor.

932 In principle, another reason for $AIC_{MIN} > AIC_{FULL} + 2$ could be that none of the models was
933 appropriate for that dataset although this has been rarely observed in relatively good datasets (Slob
934 and Setzer, 2014).

935 Finally, the results from the fitted models need to be combined to establish the final BMD confidence
936 interval. The ideal way to proceed is by model averaging (see section 2.5.6), where each of the
937 models that was fitted is taken into account, including the models that showed a less good fit. The
938 latter does not harm as model averaging uses the AIC as a weight, so that poorly fitting models will
939 hardly contribute to the final BMD confidence interval.

940 If the required model averaging software is not available, a distinction is made between models with a
941 relatively good and those with a relatively poor fit. The set of relatively good models include the
942 model with the minimum AIC and all models with an AIC no more than two units larger than that. The
943 lowest BMDL and highest BMDU from these selected models will then be used to define the BMD
944 confidence interval. It should be noted that no confidence level can be associated with this interval; in
945 general it will be larger than the nominal value of 90% used for the BMD confidence intervals
946 obtained with individual models. Hence, the BMDL will generally be smaller than the final BMDL
947 derived from model averaging. Further, it should be noted that the choice of 2 units difference
948 between AICs, as substantiated by Burnham and Anderson (2002), constitutes a somewhat arbitrary
949 way of defining the cut-off between relatively good and relatively poor models. In specific cases one
950 may decide to use a larger value than 2, for example when it would lead to the selection of just one
951 model. This problem is avoided in the approach of model averaging.

952 Before deciding to use a larger value than 2 for the AIC criterion or in situations where there is an
953 alert, the Scientific Committee recommends to consult a specialist in BMD analysis.

954

955

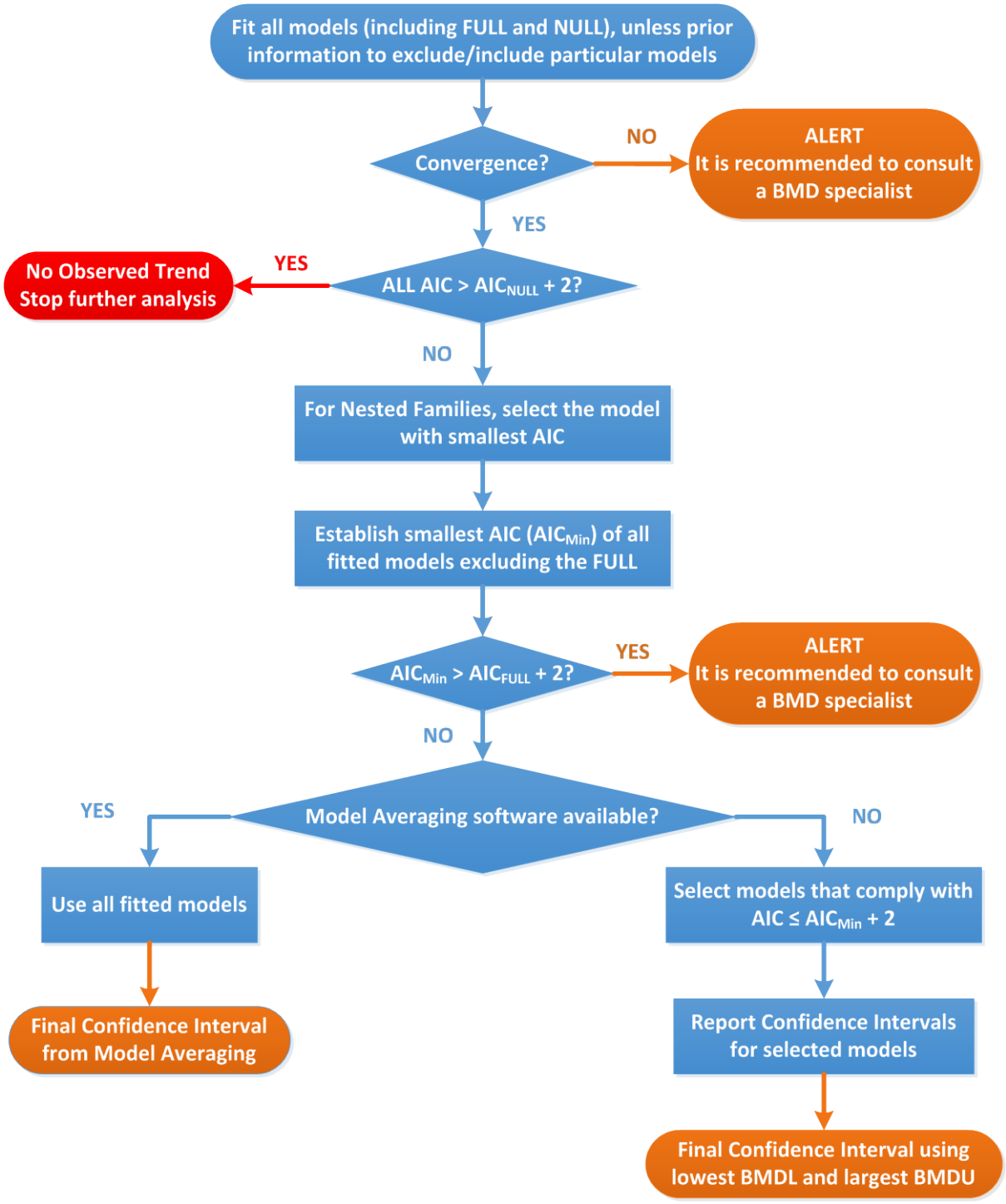


Figure 8: Flowchart to establish the BMD confidence interval and BMDL for dose-response dataset of a specified critical endpoint.

AIC = Akaike Information Criterion (indicative of the goodness of fit of the model considered)
AIC$_{NULL}$ = AIC value of the Null Model
AIC$_{FULL}$ = AIC value of the Full Model
AIC$_{MIN}$ = AIC value of the model with the lowest AIC value, the null and full models being excluded

Judging the width of the BMD confidence interval for a given dataset

Ideally, when the experimental data provide sufficient information on the dose-response relationship, the different models will result in similar confidence intervals, thereby providing an adequate basis to define a RP for the establishment of a health-based guidance value or for the calculation of a MOE (see Section 2.4).

In some cases, however, the dose response relationship may not be well defined by the data. For instance, there may be large gaps between consecutive response levels, or the lowest non-zero dose already resulted in a response much larger than the BMR. Therefore, it may occur that the applied models result in widely different BMD confidence intervals, or that some, or all of them, are very wide (several orders of magnitude). When the width of the combined BMD confidence interval is found to cover orders of magnitude, the BMDL could be orders of magnitude lower than the true BMD, had better data been available. Therefore, the resulting RP, and the HBGV or MOE eventually derived from it, might have been much higher or larger, respectively. The following options may be considered, preferably in the following order:

- Explore if there is a possibility to request for better data.
- Re-analyse the data, taking into account prior information on typical values of the shape parameters, if available from historical data, e.g. by constraining the shape parameters, or by applying prior distributions in a Bayesian approach. Whatever option is applied, this should be clearly documented. This option may be considered when the combined confidence interval is wide for various reasons related to limitations in the data, such as (i) a small total number of animals (or other experimental units) in the study, (ii) considerable scatter in the consecutive (mean) responses with increasing dose, (iii) few doses in the study design, or few doses with distinct responses, (iv) relatively small response in the top dose(s), and (v) relatively high response at the lowest dose (see previous bullet).

Determining the reference point for a given substance

The flowchart results in a final BMD confidence interval for a given dose-response dataset related to a specific endpoint. The BMD confidence interval should be derived for all datasets considered relevant (potentially leading to the RP), resulting in a set of confidence intervals indicating the uncertainty ranges around the true BMD for the endpoints considered. This set of BMD confidence intervals concisely reflects the information provided by the available data and provides the starting point for the risk assessor to derive the reference point. One way to proceed is to simply select the endpoint with the lowest BMDL and use that value as the reference point. However, this procedure may not be optimal in all cases, and the risk assessor might decide to use a more holistic approach, where all relevant aspects are taken into account, such as the BMD confidence intervals (rather than just the BMDLs), the biological meaning of the relevant endpoints, and the consequences for the HBGV or the MOE. This process will differ from case to case and it is the risk assessor's responsibility to make a substantiated decision on what BMDL will be used as the reference point. One example is a situation where the BMD confidence interval with the lowest BMDL is orders of magnitude wide. This means that the true BMD might be much higher than the BMDL, which raises the question if that BMDL would be an appropriate reference point. To answer that question, following aspects may be considered:

- If the associated HBGV would still be much higher than the exposure estimate, or the associated MOE much larger than 10'000, then the high uncertainty in the reference point, as indicated by the wide confidence interval, has no consequence for the hazard characterization. It should be however kept in mind that an exposure estimate is not a fixed value (it may well change in the future) and is therefore uncertain[11].
- In some cases, the selected reference point may not be the lowest BMDL, for example when this lowest BMDL concerns an effect that is also reflected by other endpoints (e.g. the combination of liver necrosis and serum enzymes) that resulted in much smaller confidence

[11] See http://www.efsa.europa.eu/sites/default/files/160321DraftGDUncertaintyInScientificAssessment.pdf

intervals but with higher BMDLs. In that case it might be argued that the true BMDs for those analogous endpoints would probably be similar, but one of them resulted in a much wider confidence interval (e.g. due to large measurement errors).

2.5.8. Epidemiological dose-response data

In principle, the BMD approach would also be applicable to human data. BMD analysis of human data will be the subject of a separate guidance document of the EFSA Scientific Committee.

2.5.9. Reporting of the BMD analysis

The results of a BMD analysis should be reported in such a way that others are able to follow the process.

In reporting a BMD analysis for a particular study, it is not necessary to provide information on all the endpoints analysed but only for the critical one(s) in that study. It should be made clear in a narrative why this / these endpoint(s) was / were selected.

The following information should be provided:

A. A summary table of the data for the endpoint(s) for which the BMD analysis is reported. For quantal endpoints both the number of responding animals and the total number of animals should be given for each dose level; for continuous endpoints the mean responses and the associated SDs (or SEMs) and sample sizes[12] should be given for each dose level.

B. The value of the BMR chosen, and, if deviating from the default value, the rationale for that.

C. The software used, including version number

D. Settings and statistical assumptions in the model fitting procedure when they deviate from the recommended defaults in this opinion, together with the rationale for doing so.

E. A table presenting the models used (preferably in the order of Table 3), including the null and full model and their AICs, with the BMD confidence intervals. BMDL and BMDU values should be reported with two significant figures – see examples.

F. A plot of the fitted average model. If model averaging was not used, a plot of all the models fitted to the data for the critical endpoint(s). In case of nested families, a plot of the selected model for each family

G. Conclusion regarding the selected BMDL to be used as a reference point.

A template is annexed to ensure a standardised reporting of the above-mentioned information (Appendix B).

The reporting of a BMD analysis is illustrated below for a continuous and a quantal dataset.

Whilst efforts have been made in this opinion to provide guidance on the use of BMD software, users should be aware that such software still evolves, just like the BMD approach itself. The version of the software available at the time of use may not be the same as that referred to here, but, the reporting structure should remain the same.

[12] Note that, when the individual data were used in the original analysis, slightly different results may be obtained using the summary data in the analysis.

Example 1: Continuous data.

The BMD analysis given below may serve as an example of how to report the results from a BMD analysis of a continuous dataset in an EFSA opinion. This example was run using the PROAST software (see Appendix A for an overview of the differences between PROAST and BMDS).

The data in this example relate to a 2-year study in male mice. A dose-related decrease of body weight was observed. This endpoint is assumed to be the critical effect.

A. The data

dose (mg/kg bw/day)	body weight, group mean (g)	SD	n	Sex
0	43.85	2.69	37	M
0.1	43.51	2.86	35	M
0.5	40.04	3.00	43	M
1.1	35.09	2.56	42	M

B. BMR: Default value (percent change = 5%)

C. Software used: PROAST version 61.6

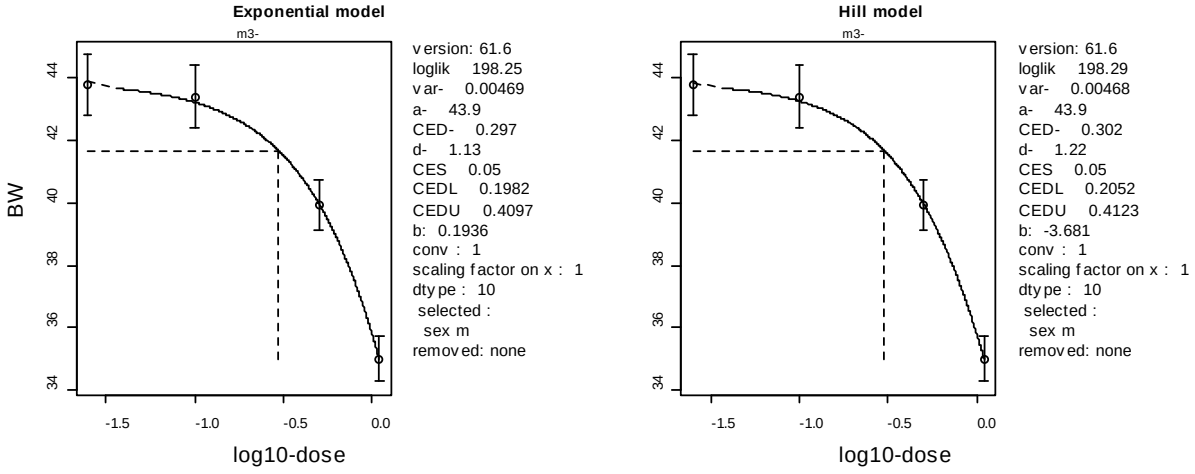
D. Additional assumptions: None

E. Table of results

Model	N° of parameters (variance excluded)	AIC		$BMDL_{05}$* (mg/kg)		$BMDU_{05}$* (mg/kg)	
		Exponential	Hill	Exponential	Hill	Exponential	Hill
Null model	1	-234.06					
Model 3	3	-388.50*	-388.58*	**0.198**	**0.205**	**0.410**	**0.412**
Model 5	4	-386.72	-386.72				
Full model	4	-386.72					

* Selected model, based on lowest AIC.

F. Figure of fitted models



Fitted curves for model 3 from the exponential model family (left panel) and model 3 from the Hill model family (right panel). Vertical whiskers represent 95%-confidence intervals for the responses. Dose is plotted on log-scale for better readability; the response in the controls is shown at an arbitrary level lower than the lowest non-zero dose (as zero dose is situated at minus infinity on log-scale).

G. Conclusion

There were no alerts (the fit was reached under convergence, and the AICs of both models differed less than two units from the full model).

For both the exponential and the Hill family of models, model 3 was selected, based on the lowest AIC. The two associated BMD confidence intervals were similar. Therefore, model averaging would hardly provide a different result, and it was decided to select the lowest BMDL and highest BMDU from both models (in this case, they were the same for both models when using two significant figures).

The combined BMD confidence interval was (0.20, 0.41) mg/kg.

The $BMDL_{05}$ for this dataset is 0.20 mg/kg.

1085 **Example 2: quantal data**

1086 This example relates to a 2-year study in rats, where three doses of a substance were administered to
1087 the animals. Dose-related changes in thyroid epithelial cell vacuolization were found, and these data
1088 were used for a BMD analysis. The BMD analysis given below may serve as an example of how to
1089 report the results from a BMD analysis of a quantal dataset in an EFSA opinion.

1090

1091 **A. The data**

Dose (mg/kg-day)	N° of animals with thyroid epithelial vacuolization	N° of animals in dose group	Sex
0	6	50	f
3	6	50	f
12	34	50	f
30	42	50	f

1092

1093 **B. BMR:** Default value (extra risk = 10%)

1094 **C. Software used:** MADr-BMD program (Wheeler and Bailer, 2008) + PROAST version 61.6

1095 **D. Additional assumptions:** None

1096 **E. Table of results**

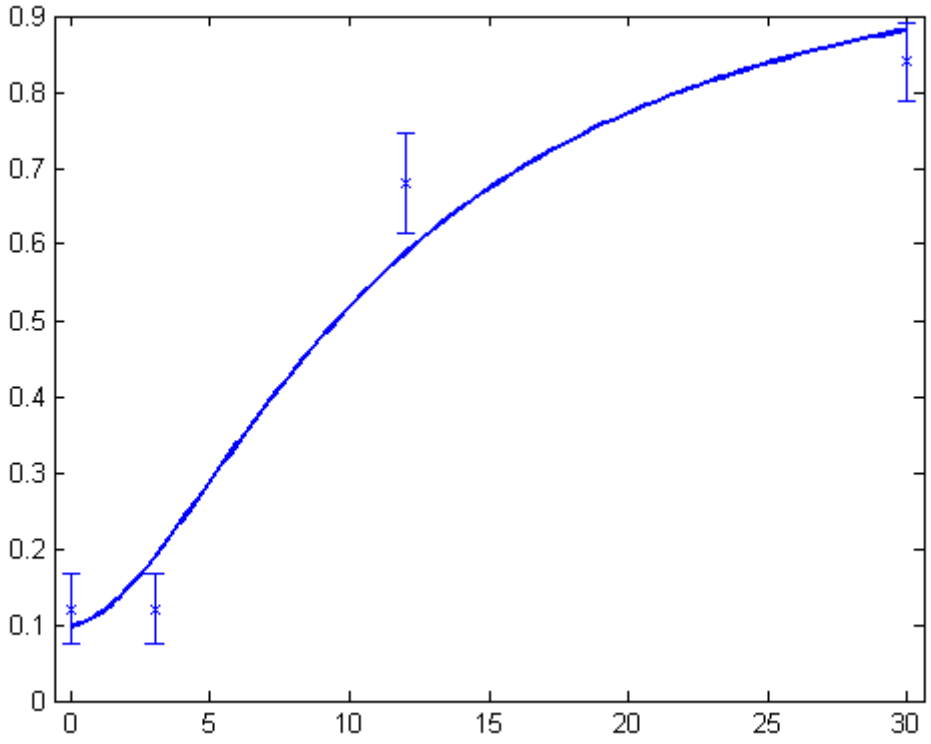
Model	N° of parameters	AIC	$BMDL_{10}$	$BMDU_{10}$[c]
Null Model	1	276.38	--	--
Gamma	3	192.99	1.21	2.67[c]
Logistic	2	198.47	3.31	4.90[c]
LogLogistic	**3**	**189.81**[b]	**1.84**	**5.00**[c]
Probit	2	199.07	3.37	NA
LogProbit	**3**	**189.73**[a]	**1.98**	**5.11**[c]
Weibull	3	193.55	1.10	4.01[c]
LMS (Two stage)	3	194.20	1.35	3.10
Full Model	4	188.04	--	--

[a] Model with lowest AIC
[b] AIC differs less than two units from lowest AIC
[c] Calculated by PROAST, as BMDS does not yet provide BMDUs except for the two-stage model.

1097

1098

1099 **F. Figure of fitted model**



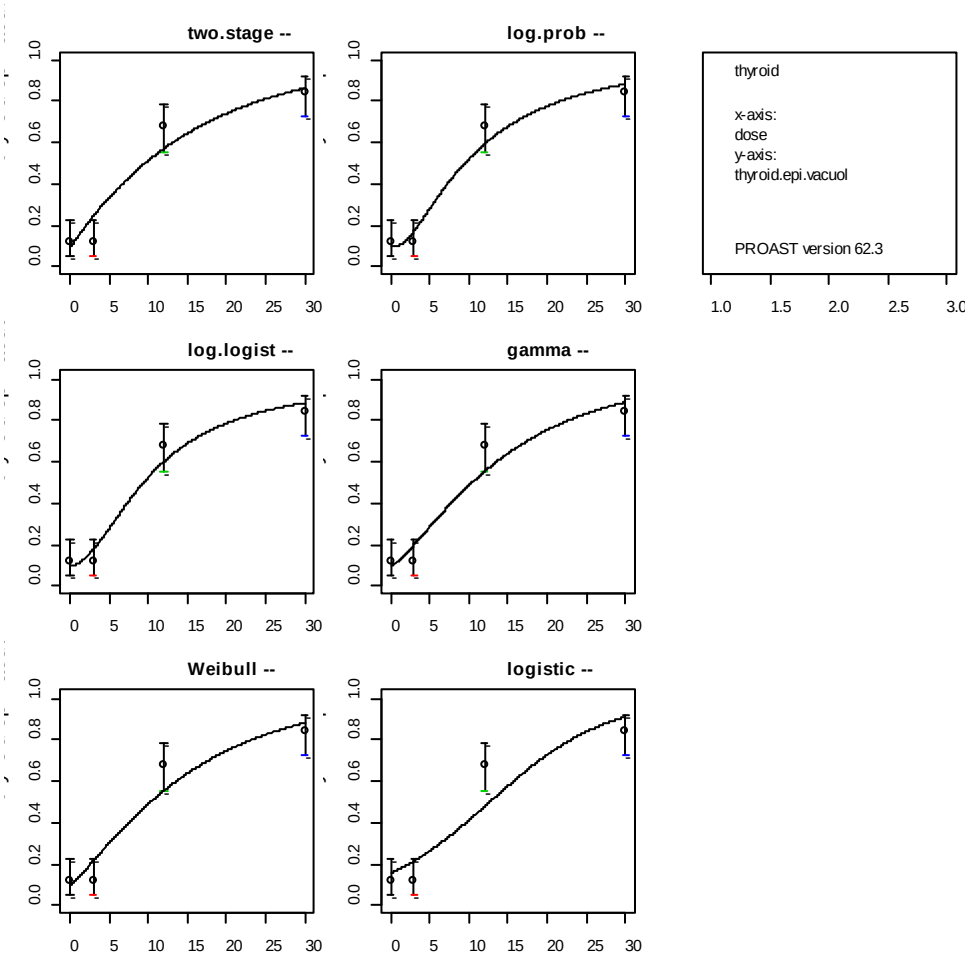
1100 "Average model" from model averaging analysis of the observed incidences of animals with thyroid
1101 epithelial cell vacuolization. The average model was constructed via averaging all weighted-model
1102 results at a finite set of points (i.e., doses) in order to generate curve. The MADr-BMD program
1103 (Wheeler and Bailer, 2008) was used.

1104

1105 If model averaging software was not available, the plots of the recommended models should be
1106 shown:

1107

1108



The recommended models fitted to the observed incidences of animals with thyroid epithelial cell vacuolization, with 95% confidence intervals at each response. PROAST version 62.3 was used.

G. Conclusion

There were no alerts (the fit was reached under convergence, and the AIC differed less than two units from the full model).

The preferred way of combining the results is by model averaging. The MADr-BMD program as described in Wheeler and Bailer (2008) can be used for that purpose. In this approach all models are taken into account with a weight that is derived from the AIC. The following weights were used for this dataset:

Model	Weight	AIC
Log-probit	0.411	189.73
Log-logistic	0.395	189.81
Gamma	0.080	192.99
Weibull	0.061	193.54
Multistage 2°	0.044	194.20
Logistic	0.005	198.47
Probit	0.004	199.07

The current MADr-BMD program does not calculate the BMDU, it only calculates a BMDL (and a BMD point estimate). The BMDL for this dataset was found to be 1.5 mg/kg based on model averaging.

If model averaging software is not available, the surrogate method may be used, where the lowest BMDL and highest BMDU is taken from the models that showed an AIC differing less than 2 units from the lowest AIC. In this dataset only the log-probit and the log-logistic models meet that criterion. Combining these two models results in a BMD confidence interval of (1.8, 5.1) mg/kg.

The fact that the model averaging software resulted in a slightly lower BMDL is due to the fact that the other models were taken into account as well (although with low weight).

3. Conclusions

This revised guidance takes account of the experience accumulated in BMD analysis over the last seven years.

The Scientific Committee confirms that the BMD approach is a scientifically more advanced method compared to the NOAEL approach for deriving a Reference Point, since it makes extended use of dose-response data and it provides a quantification of the uncertainty in the estimated Reference Point resulting from the statistical limitations in the dose-response data. Using the BMD approach results in a more consistent RP, as a consequence of the specified BMR. Health-based guidance values derived using the BMD approach can be expected to be as protective as those derived from the NOAEL approach, i.e. on average over a large number of risk assessments. Therefore the default values for uncertainty factors currently applied are equally applicable.

The Scientific Committee does not consider it necessary to repeat all previous evaluations using the NOAEL approach by the BMD approach, because, on average, the two approaches give comparable results. Similarly, the SC does not consider it necessary to repeat previous risk assessments related to quantal endpoints that used the 2009 version of the BMD guidance, given the modifications proposed in the updated version of the guidance for this type of data. Regarding previous risk assessments where the 2009 BMD guidance was applied to continuous datasets, the updated guidance might result in lower reference points, in particular when model 2 of the nested families was selected to derive the Reference Point.

Where refinement of previous risk assessments is considered necessary, the BMD approach as described in this guidance should be applied.

The BMD approach is applicable to all chemicals in food, independently of their category or origin, e.g. pesticides, additives or contaminants, for identifying reference points to establish health-based guidance values or to calculate Margins of Exposure. The BMD approach can be used for dose response assessment of experimental animal data as well as for epidemiological data, although the latter is not addressed in this guidance document and will be subject to a separate guidance of the EFSA Scientific Committee.

4. Recommendations:

- The Scientific Committee strongly recommends that the BMD approach, and more specifically model averaging, is used for the determination of the Reference Points for establishing health-based guidance values and for calculating margins of exposure. To that end, the Scientific Committee recommends that training in dose-response modelling and the use of BMD software continues to be offered to experts in the scientific Panels and EFSA Units.
- The Scientific Committee is firmly of the view that current toxicity test guidelines should be reconsidered given the expected wide application of the BMD approach, e.g. increase the number of dose levels without changing the total number of animals used in the experiment.
- The Scientific Committee recommends EFSA to establish a BMD Standing Working Group to be consulted by EFSA experts and staff on BMD analysis issues if needed, e.g. when alerts are

identified or when applying the BMD approach to histopathological (ordinal) data. A network on BMD, coordinated by EFSA, should also be considered to exchange experience and develop expertise with EFSA Partners (Member States competent, EU sister agencies, DG Santé Scientific Committees and international organisations).

- The Scientific Committee identified the need for a specific guidance on the use of the BMD approach to analyse human data.

References

Acquah H de-G, 2010. Comparison of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in Selection of an Asymmetric Price Relationship. Journal of Development and Agricultural Economics 2, 1-6.

Allen BC, Kavlock, RJ, Kimmel CA and Faustman EM, 1994. Dose-response assessment for developmental toxicity. II. Comparison of generic Benchmark dose estimates with No Observed Adverse Effects Levels. Fundamental and Applied Toxicology, 23, 487-495.

Baird SJS, Cohen JT, Graham JD, Shylakter AI and Evans JS, 1996. Noncancer risk assessment: A probabilistic alternative to current practice. Human and Ecological Risk Assessment, 2, 79–102.

Bemis JC, Wills JW, Bryce SM, Torous DK, Dertinger SD and Slob W, 2015. Comparison of In Vitro and In Vivo Clastogenic Potency Based on Benchmark Dose Analysis of Flow Cytometric Micronucleus Data. Mutagenesis, gev041. doi:10.1093/mutage/gev041

Bokkers BGH and Slob W, 2005. A Comparison of Ratio Distributions Based on the NOAEL and the Benchmark Approach for Subchronic-to-Chronic Extrapolation. Toxicological Sciences, 85, 1033-1040

Bokkers BGH and Slob W 2007. Deriving a data-based interspecies assessment factor using the NOAEL and the Benchmark dose approach. Critical Review in Toxicology Journal, 37, 353-377.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D and Farland W, 2006. IPCS Framework for analysing the relevance of a cancer mode of action for humans. Critical Review in Toxicology Journal, 36, 781-792.

Bosgra S, van der Voet H, Boon PE and Slob W, 2009. An integrated probabilistic framework for cumulative risk assessment of common mechanism chemicals in food: An example with organophosphorus pesticides. Regulatory Toxicology and Pharmacology Journal, 54, 124-133.

Budtz-Jørgensen E, Keiding N and Grandjean P, 2001. Benchmark Dose Calculation from Epidemiological Data. Biometrics 57, 698-706.

Budtz-Jørgensen E, Keiding N and Grandjean P, 2004. Effects of exposure imprecision on estimation of the benchmark dose. Risk Analysis, 24, 1689-1696.

Budtz-Jørgensen E, 2007. Estimation of the benchmark dose by structural equation models. Biostatistics, 8, 4, 675-688.

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

Burnham KP and Anderson DR, 1998. Model selection and multimodel inference. A practical information – Theoretic approach. Second edition 2002. Springer-Verlag, New York, USA.

Chiu WA and Slob W, 2015. A Unified Probabilistic Framework for Dose-Response Assessment of Human Health Effects. Environmental Health Perspectives (in press). http://ehp.niehs.nih.gov/1409385/

Claeskens G and Hjort NL, 2008. Model Selection and Model Averaging. Cambridge University Press, UK.

Crump KS 1984. A New Method for Determining Allowable Daily Intakes. Fundamental and Applied Toxicology, 4, 854-871.

Crump K 2002. Critical issues in benchmark calculations from continuous data. Critical Review in Toxicology Journal, 32, 133-153.

EFSA (European Food Safety Authority), 2005. Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. The EFSA Journal 2005, 282, 1-31

1225 EFSA (European Food Safety Authority), 2007. Opinion of the scientific panel on contaminants in the
1226 food chain [CONTAM] related to the potential increase of consumer health risk by a possible
1227 increase of the existing maximum levels for aflatoxins in almonds, hazelnuts and pistachios and
1228 derived products. The EFSA Journal 446, 1-127

1229 EFSA (European Food Safety Authority), 2008. Opinion of the Scientific Panel on Plant Protection
1230 products and their Residues to evaluate the suitability of existing methodologies and, if
1231 appropriate, the identification of new approaches to assess cumulative and synergistic risks from
1232 pesticides to human health with a view to set MRLs for those pesticides in the frame of Regulation
1233 (EC) 396/2005. The EFSA Journal 2008, 704, 1-84

1234 EFSA (European Food Safety Authority), 2009. Guidance of the Scientific Committee on a request from
1235 EFSA on the use of the benchmark dose approach in risk assessment. EFSA Journal 2009, 1150, 1-
1236 72.

1237 EFSA (European Food Safety Authority), 2009a. Safety evaluation of ractopamine. Scientific Opinion of
1238 the Panel on Additives and Products or Substances used in Animal Feed. EFSA Journal 2009, 1041,
1239 1-52.

1240 EFSA (European Food Safety Authority), 2009b. Cadmium in food. Scientific Opinion of the Panel on
1241 Contaminants in the Food Chain. EFSA Journal, 2009, 980, 1-139.

1242 EPA (environmental Protection Agency), 1995. The use of the benchmark dose approach in health risk
1243 assessment. EPA/630/R-94/007. Risk Assessment Forum, Washington DC.

1244 FOSIE (Food Safety in Europe) 2002. Risk assessment of chemicals in food and diet. Food and
1245 Chemical Toxicology, 40, 2-3.

1246 Fowles JR, Alexeeff GV and Dodge D, 1999. The use of benchmark dose methodology with acute
1247 inhalation lethality data. Regulatory Toxicology and Pharmacology 29, 262-278.

1248 Fryer M, Collins CD, Ferrier H, Colvile RN and Nieuwenhuijsen MJ, 2006. Human exposure modeling
1249 for chemical risk assessment: A review of current approaches and research and policy implications.
1250 Environmental Science & Policy, 9, 261–274.

1251 Gibney MJ and van der Voet H, 2003. Introduction to the Monte Carlo project and the approach to the
1252 validation of probabilistic models of dietary exposure to selected food chemicals. Food Additives
1253 and Contaminants, 20 (suppl. 1), S1–S7.

1254 IPCS (International Program on Chemical Safety), 2014. Guidance Document on Evaluating and
1255 Expressing Uncertainty in Hazard Characterization. World Health Organization, Geneva. Available
1256 at http://www.who.int/ipcs/methods/harmonization/areas/hazard_assessment/en/ [accessed 28
1257 April 2015]

1258 Joint FAO (Food and Agriculture Organization of the United Nations) and WHO (World Health
1259 Organization). Expert committee on food Additives – JECFA (2006a), Sixty-fourth meeting,
1260 WHO/IPCS Safety evaluation of certain contaminants in food. WHO Food Additives Series 55.

1261 Joint FAO (Food and Agriculture Organization of the United Nations) and WHO (World Health
1262 Organization). Expert committee on food Additives – JECFA (2006b). WHO Technical Report Series
1263 939. Evaluation of certain veterinary drug residues in food. 66th report of the Joint FAO/WHO
1264 Expert Committee on food additives.

1265 Kang SH, Kodell RL and Chen JJ, 2000. Incorporating model uncertainties along with data
1266 uncertainties in microbial risk assessment. Regulatory Toxicology and Pharmacology, 32, 68-72.

1267 Kavlock RJ, Allen BC, Faustman EM and Kimmel CA, 1995. Dose-response assessments for
1268 developmental toxicity IV. Benchmark doses for fetal weight changes. Fundamental and Applied
1269 Toxicology, 26, 211-222.

1270 Kienhuis, AS, Slob W, Gremmer ER, Vermeulen JP and Ezendam J, 2015. A dose-response modelling
1271 approach shows that effects from mixture exposure to the skin sensitizers are in line with dose
1272 addition and not with synergism. Toxicological Sciences, kfv109. doi: 10.1093/toxsci/kfv109

1273 Murata K, Budtz-Jorgensen E and Grandjean P, 2002. Benchmark dose calculations for
1274 methylmercury-associated delays on evoked potential latencies in children. Risk Analysis, 22, 465-
1275 74.

1276 Piersma AH, Verhoef A, te Biesebeek JD, Pieters MN and Slob W, 2000. Developmental toxicity of
1277 butyl benzyl phthalate in the rat using a multiple dose study design. Reproductive Toxicology, 14
1278 (5), 417-425.

1279 Pieters MN, Bakker M and Slob W, 2004. Reduced intake of deoxynivalenol in The Netherlands: a risk
1280 assessment update. Toxicology Letters, 153, 145-153.

1281 Sand S, Falk Filipsson A and Victorin K, 2002. Evaluation of the benchmark dose method for
1282 dichotomous data: model dependence and model selection.
1283 Regulatory Toxicology and Pharmacology, 36, 184-197.

1284 Sand S, von Rosen D, Victorin K and Falk Filipsson A, 2006. Identification of a critical dose level for
1285 risk assessment: developments in benchmark dose analysis of continuous endpoints. Toxicological
1286 Sciences, 90, 241-251.

1287 Sand S, Portier CJ and Krewski D, 2011. A Signal-to-Noise Crossover Dose as the Point of Departure
1288 for Health Risk Assessment. Environmental Health Perspectives, 119(12), 1766-1774

1289 Shao K, Gift JS and Setzer RW, 2013. Is the assumption of normality or log-normality for continuous
1290 response data critical for benchmark dose estimation?. Toxicology and applied pharmacology,
1291 272(3), 767-779.

1292 Slob W, 1994. Uncertainty Analysis in Multiplicative Models. Risk Analysis, 14: 571–576.

1293 Slob W and Pieters MN, 1998. A probabilistic approach for deriving acceptable human intake limits and
1294 human health risks from toxicological studies: general framework. Risk Analysis 18, 787-798.

1295 Slob W, 2002. Dose-response modelling of continuous endpoints. Toxicological Sciences, 66 (2), 298-
1296 312.

1297 Slob W, Bakker MI, Biesebeek JDT and Bokkers BG, 2014. Exploring the Uncertainties in Cancer Risk
1298 Assessment Using the Integrated Probabilistic Risk Assessment (IPRA) Approach. Risk Analysis,
1299 34(8), 1401-1422.

1300 Slob W, 2014. Benchmark dose and the three Rs. Part II. Reduction by getting the same information
1301 from fewer animals. Critical Reviews in Toxicology 44, 568-580.

1302 Soeteman-Hernández LG, Johnson GE and Slob W 2015a. Estimating the carcinogenic potency of
1303 chemicals from the in vivo micronucleus test. Mutagenesis, gev043, doi:10.1093/mutage/gev043

1304 Soeteman-Hernández LG, Fellows MD, Johnson GE and Slob W, 2015b. Correlation of in vivo versus in
1305 vitro Benchmark doses (BMDs) derived from micronucleus test data: A proof of concept study.
1306 Toxicol Sci, kfv189.

1307 Swartout JC, Price PS, Dourson ML, Carlson-Lynch HL and Keenan RE, 1998. A Probabilistic
1308 Framework for the Reference Dose (Probabilistic RfD). Risk Analysis**,** 18, 271-282

1309 Tressou J, Leblanc JC, Feinberg M and Bertail P, 2004. Statistical methodology to evaluate food
1310 exposure to a contaminant and influence of sanitary limits: Application to Ochratoxin A. Regulatory
1311 Toxicology and Pharmacology, 40, 252–263.

1312 United States Environmental Protection Agency, 2016. Categorical Regression (CATREG) User Guide
1313 (Version 3.0.1.5). Available at: https://www.epa.gov/sites/production/files/2016-
1314 03/documents/catreg_user_guide.pdf

1315 United States Environmental Protection Agency (US EPA), 2012. Benchmark Dose Technical Guidance.
1316 (EPA/100/R-12/001). Washington DC: Risk Assessment
1317 Forum. http://www.epa.gov/raf/publications/pdfs/benchmark_dose_guidance.pdf

1318 Van der Voet H and Slob W, 2007. Integration of probabilistic exposure assessment and probabilistic
1319 hazard characterization. Risk Analysis, 27, 351-371.

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(5), 1 – 15.

WHO (World Health Organization), 1987. Principles for the Safety Assessment of Food Additives and Contaminants in Food. Environmental Health Criteria 70, WHO/IPCS.

Wills JW, Johnson GE, Doak SH, Soeteman-Hernández LG, Slob W, White PA, 2015. Empirical Analysis of BMD Metrics in Genetic Toxicology Part I: In Vitro Analyses to Provide Robust Potency Rankings. Mutagenesis, gev085.

Woutersen RA, Jonker D, Stevenson H, te Biesebeek JD and Slob W, 2001. The BMD approach applied to a 28-day toxicity study with Rhodorsil Silane in rats: the impact of increasing the number of dose groups. Food and Chemical Toxicology, 39, 697-707.

Abbreviations

ADI	Acceptable Daily Intake
AIC	Akaike Information Criterion
BMD	Benchmark Dose
BMDL	Lower confidence limit of the benchmark dose (equivalent term: CEDL)
BMDU	Upper confidence limit of the benchmark dose (equivalent term: CEDU)
BMR	Benchmark Response
CEDL	See BMDL
CEDU	See BMDU
FAO	Food and Agriculture Organization of the United Nations
GUI	Graphical User Interface
HBGV	Health-Based Guidance Value
IPCS	WHO International Programme on Chemical Safety
JECFA	Joint FAO/WHO Expert Committee on Food Additives
JMPR	Joint FAO/WHO Meeting on Pesticide Residues
LOAEL	Lowest-Observed-Adverse-Effect-Level
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
SD	Standard Deviation
SEM	Standard Error of the Mean
TDI	Tolerable Daily Intake
TEF	Toxic Equivalency Factor
WHO	World Health Organization

Appendix A – Summary of the differences between BMDS and PROAST

A summary of the main differences between BMDS and PROAST software is provided in Table 4.

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) in a specific dose-response analysis, this should result in only slight differences in the BMD and BMDL when the within-group variation is small (Shao et al. 2013); larger deviations may occur with data showing large within-group variation. In view of this guidance document it is important to note that BMDS lacks the option of setting the distribution to lognormal in the Hill model.

The procedure in PROAST for fitting the family of exponential models is available in the BMDS, but the family of Hill models cannot be fitted in BMDS. In view of this guidance document, an important gap in BMDS is that the three-parameter Hill model cannot be fitted.

It should also be noted that the parameterization of the Hill model differs between BMDS and PROAST, but the models may still be regarded as equivalents. Besides the four-parameter Hill model and the nested family of exponential models, BMDS also includes 3 additional models (power, linear and polynomial) for continuous data. These models are, however, not recommended in this opinion. Table 5 lists the models used in BMDS that are equivalent to those recommended in Table 3.

For continuous data, the variance can be either specified as constant or it can be modelled as a function of the mean response in BMDS, while PROAST always assumes the variance to be constant on log-scale (i.e. 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, four ways of defining the BMR are available for continuous data: standard deviation (Std. Dev), relative deviation (Rel. Dev), absolute deviation (Abs. Dev), and Point. 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 all models. In view of this guidance document recommending the consideration of the BMD confidence interval, the lack of BMDUs in the BMDS output is a limitation.

For analysis of quantal data, BMDS and PROAST are essentially the same. Similar to the case for the Hill model, however, many of the quantal models differ in parameterisation between BMDS and PROAST, but they do provide similar results.

1368 **Table 4:** 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
BMDU calculated	No, except for Multistage Cancer model	Yes
Default assumption of distribution continuous data	normal	lognormal
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	No
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 statistical model assumptions	No	Yes, by residual plots

Evaluation of dose addition	No	Yes

Table 5: Dose-response models for continuous and quantal data in BMDS (EPA, 2012).

Continuous models	
Hill model $\mu(X) = \gamma + \nu \frac{X^{\eta}}{\kappa^{\eta} + X^{\eta}}$	exponential models (a set of nested models) Model 3: $\mu(X) = \gamma + e^{(kX)^d}$ Model 5: $\mu(X) = \gamma\left(c - (c-1)e^{-(kX)^d}\right)$
Quantal models	
logistic model $p(X) = \frac{1}{1+e^{-(\alpha+\beta X)}}$ log-logistic model $p(X) = \gamma + \frac{1-\gamma}{1+e^{-(\alpha+\beta \ln X)}}$ probit model[a] $p(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta X} e^{\frac{-x^2}{2}} dx$ log-probit model[a] $p(X) = \gamma + \frac{1-\gamma}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta \ln X} e^{\frac{-x^2}{2}} dx$	Weibull model $p(X) = \gamma + (1-\gamma)\left(1 - e^{-\beta(X)^{\alpha}}\right)$ gamma model[b] $p(X) = \gamma + (1-\gamma)\frac{1}{\Gamma(\alpha)} \int_0^{\beta X} x^{(\alpha-1)} e^{-x} dx$ multi-stage model $p(X) = \gamma + (1-\gamma)\left(1 - e^{-\sum_{j=1}^{n} \beta_j X^j}\right)$

For continuous models the variance across dose group may either be assumed to be constant or non-constant (a power function of the mean response).

[a] In the model, $\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ is the standard normal density function.

[b] In the model, $\Gamma(\alpha) = \int_0^{\infty} x^{(\alpha-1)} e^{-x} dx$ is the gamma function.

Appendix B – Template for reporting a BMD analysis

A Data description

Brief general description of the data. This section should include a table summarizing the data. In case that raw data is obtained/provided, resulting in a too large table, summary statistics may be given instead[13]. For quantal endpoints both the number of responding animals and the total number of animals should be given for each dose level; for continuous endpoints either the individual responses or the mean responses and the associated SDs (or SEMs) and sample sizes should be given for each dose level.

Table 6: Example of table for continuous dose-response data

Dose	Endpoint mean	SD	N	Covariates (gender)
0	43.85	2.69	37	M
0.1	43.51	2.86	35	M
0.5	40.04	3.00	43	M
1.1	35.09	2.56	42	M
0	41.54	6.26	36	F
0.1	38.71	4.73	42	F
0.5	33.76	3.92	37	F
1.1	28.55	2.08	38	F

In case that several control groups are reported in the publication or provided by the applicant, they should all be presented in the table.

Table 7: Example of table for quantal dose-response data

Dose	Number of animals with event of interest	N	Covariates (gender)
0	2	50	M
3	4	50	M
12	32	49	M
30	45	50	M
0	6	50	F
3	6	50	F
12	34	50	F
30	42	50	F

In case different endpoints are to be analysed, they should be described in different subsections, containing information pertaining to each endpoint.

The following steps (until section 3) apply for each endpoint considered.

[13] Note that, when the individual data were used in the original analysis, slightly different results may be obtained using the summary data in the analysis.

B Selection of the BMR

Specification of a low and measurable response level or decrease in response compared with the background response, called BMR. The rationale behind the choice made should be described, in particular when it deviates from the default.

C Software used

The software used, including version number should be reported. In case a pre-existing software was not used, the script for the BMD analysis should be provided as an appendix.

D Specification of deviations from default assumptions

- In case model averaging software is available and another approach was used, rationale for deviating from the recommended approach should be provided.
- Assumptions made when deviating from the recommended defaults in this guidance document (e.g. gamma distributional assumption instead of log-normal, heteroscedasticity instead of homoscedasticity).
- Other models than the recommended ones listed in Table 3 of this guidance document that were fitted should be listed, with the respective description of reasons to include them.
- Description of any deviation from the procedure described in the flow chart (Fig. 8) to obtain the final BMD confidence interval (e.g. using AIC+5 instead of AIC+2 for model selection).

E Results

The results of the BMD analysis should contain:

- A table presenting results of the models fitted, including number of parameters in the model, AIC, BMDL and BMDU (see Table 8)
- Report whenever convergence issues were encountered
- Report whenever the full model performed better than any of the fitted models according to the criterion $AIC_{Min} > AIC_{Full}+2$
- Highlight the models complying with the rule $AIC \leq AIC_{Min}+2$.

Table 8: Result table.

Model	Number of parameters	Log-likelihood	AIC	BMDL	BMDU
Null	1				
Full					
Exp Model 3					
Exp Model 5					
Hill Model 3					
Hill Model 5					

F Plots of fitted models

In case model averaging is used, show the plot of the data with confidence intervals for the responses, together with the resulting model average fit. If no model averaging software is available,

or the decision was made to deviate from the model averaging recommendation, show the plot with all the models fitted (in case of nested model families, the plot of the selected model for each family).

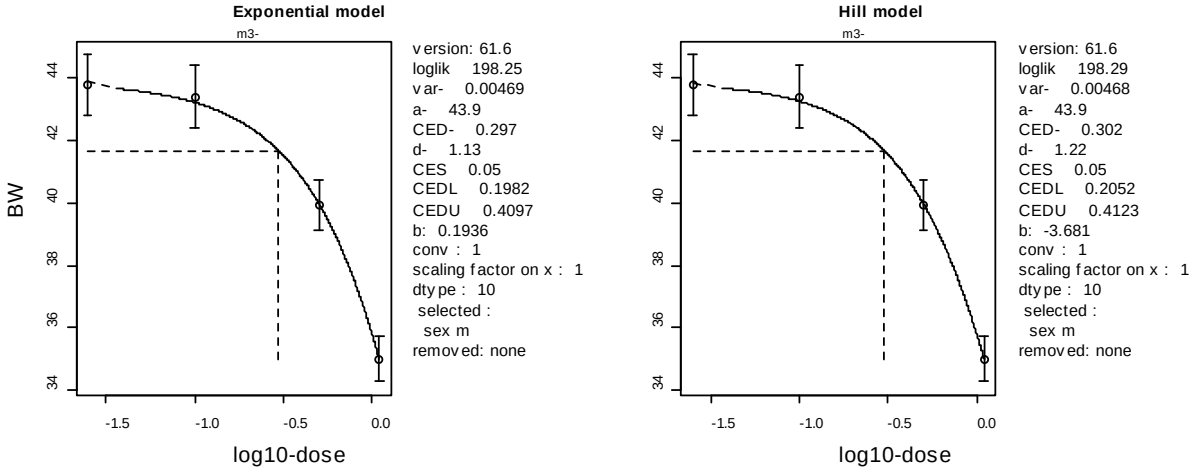


Figure 9: Plot of the selected models from each model family.

G Conclusions

The section should discuss the results for the different endpoints and the rationale to decide on the critical one on which the assessment will be based.

- Discuss if there were any alerts, and if so, how they well dealt with.
- In case model averaging was used, provide the weights for each model as used in the model averaging.
- Discuss any particular circumstances, if relevant for the final outcome of the BMD confidence interval.

The BMD confidence interval of the critical endpoint (and the BMDL selected as reference point) should be reported and discussed.