

SCIENTIFIC OPINION

Use of the benchmark dose approach in risk assessment¹

Guidance of the Scientific Committee

(Question No EFSA-Q-2005-232)

Adopted on 26 May 2009

PANEL MEMBERS

Susan Barlow, Andrew Chesson, John D. Collins, Albert Flynn, Anthony Hardy, Klaus-Dieter Jany, Ada Knaap, Harry Kuiper, John-Christian Larsen, David Lovell², Pierre Le Neindre, Jan Schans, Josef Schlatter, Vittorio Silano, Staffan Skerfving, Philippe Vannier.

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 by EFSA to assess the existing information on the utility of the benchmark dose (BMD) approach, as an alternative to the traditionally used NOAEL approach, and to make recommendations on whether EFSA should use the BMD approach and under which circumstances this use would be appropriate. The Scientific Committee was also asked to provide some guidance on how to use the BMD approach for analysing dose-response data from experimental studies, and to look at the possible application of this approach to data from observational epidemiological studies. Finally, the Scientific Committee was asked to advise on whether the selection of appropriate uncertainty factors are needed when using the BMD approach for deriving the Reference Point.

¹ For citation purposes: Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. *The EFSA Journal* (2009) 1150, 1-72

² External expert of the EFSA Scientific Committee

Traditionally, when experimental animal data are used for risk assessment of substances in food, which are not genotoxic and carcinogenic, the No-Observed-Adverse-Effect-Level (NOAEL) and/or the Lowest-Observed-Adverse-Effect-Level (LOAEL) for the critical effect of a substance, forms the Reference Point for deriving health-based guidance values, such as an Acceptable Daily Intake (ADI). However, while this approach may utilise qualitative information, it does not use the data available in a quantitative way. In contrast, the BMD approach makes extended use of the dose-response data from studies in experimental animals or from observational epidemiological studies to better characterise and quantify potential risks.

EFSA has not systematically used the BMD approach so far, although some EFSA Scientific Panels have been applying the BMD approach occasionally. Moreover, experimental data from which a BMD has been calculated are submitted from time to time. EFSA's Scientific Committee has also proposed in a previous Opinion to use the BMD approach for deriving the Reference Point for the estimation of Margins of Exposure for substances that are both genotoxic and carcinogenic.

After comparing the strengths and weaknesses of the BMD and NOAEL approaches for deriving Reference Points for risk assessment, the Scientific Committee concludes that the BMD approach is a scientifically more advanced method to the NOAEL approach for deriving a Reference Point, since it makes extended use of available dose-response data and it provides a quantification of the uncertainties in the dose-response data. Using the BMD approach also results in a more consistent Reference Point, as a consequence of the specified benchmark response.

Whilst software is currently available for BMD analysis, the Scientific Committee notes that this is evolving and significant developments are anticipated in the medium term, e.g. model averaging, analysis of continuous data.

The Scientific Committee also considers that when toxicity test guidelines, e.g. OECD guidelines, are revised, specific aspects of the BMD approach are taken into account.

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 remain appropriate and there is no need for any additional uncertainty factor.

The Scientific Committee does not consider it necessary to repeat all previous evaluations using the BMD approach, because, on average, the BMD and NOAEL approaches give comparable results. Where refinement of previous risk assessments is considered necessary, for instance where the human exposure is close to the ADI, application of the BMD approach would be of particular value.

The BMD approach is applicable to all chemicals in food, irrespective of their category or origin, e.g. pesticides, additives or contaminants. The BMD approach is of particular value for i) situations where the identification of a NOAEL is uncertain, ii) providing a Reference Point for the Margin of Exposure in case of substances that are both genotoxic and carcinogenic, and iii) dose-response assessment of observational epidemiological data. In the short term, the EFSA Scientific Panels and Units are strongly encouraged to adopt the BMD approach to situations such as those described above.

In the longer term, the Scientific Committee anticipates that the BMD approach will be used as the method of choice for the determination of the Reference Points for deriving health-based guidance values and margins of exposure. To that end, recognising that there are practical considerations regarding its introduction and wider use in EFSA, and recognising that its application requires a level of expert judgement and modelling expertise, the Scientific Committee proposes that training in dose-response modelling and the use of the software be offered to experts in the scientific Panels and EFSA Units. The Scientific Committee would then like to review the implementation, experience and acceptability of the BMD approach in EFSA's work in two years time.

Key words: Benchmark dose, BMD, BMDL, BMR, NOAEL, dose-response model, hazard characterisation, risk assessment, human data, BMDS, PROAST

TABLE OF CONTENTS

Panel Members.....	1
Summary	1
Table of Contents	4
Background as provided by EFSA	5
Terms of reference as provided by EFSA	6
Acknowledgements	6
Assessment.....	7
1. Introduction	7
2. Hazard identification: selection of potential critical endpoints	8
3. Using dose-response data in hazard characterisation	8
3.1. The NOAEL approach	9
3.2. The BMD approach	10
3.3. Interpretation and properties of the NOAEL and the BMD	12
3.4. NOAEL and BMD approach: some illustrations.....	14
3.5. Selection of the benchmark response (BMR)	21
4. Consequences for hazard/risk characterisation and risk communication.....	22
4.1. Setting health-based guidance values	22
4.2. Risk assessment of substances which are both genotoxic and carcinogenic	23
4.3. Potency comparisons	23
4.4. Probabilistic risk assessment	24
4.5. BMD vs. NOAEL: Perception of safety	24
5. Guidance to applying the BMD approach	25
5.1. Specification of type of dose-response data	26
5.2. Specification of BMR	26
5.3. Selection of candidate dose-response models.....	27
5.4. Fitting and accepting models	30
5.5. Estimating the BMD, and establishing the BMDL as the Reference Point	32
5.6. Illustrations	33
5.7. Specific issues of human dose-response data	37
5.8. Reporting of the BMD analysis	37
Conclusions and Recommendations.....	42
References	44
Appendix: Software for BMD analysis.....	47

BACKGROUND AS PROVIDED BY EFSA

Traditionally, when animal data have to be used for risk assessment of chemicals in food, the No-Observed-Adverse-Effect-Level (NOAEL) and/or the Lowest-Observed-Adverse-Effect-Level (LOAEL) for the critical effect of a substance, forms the reference point for the risk assessment. However, this approach does not take into account directly the shape of the dose-response curve, thus all available information is not utilised. The numerical NOAEL/LOAEL is also critically dependent on the choice of dose intervals made, the number of subjects at the chosen doses, as well as the sensitivity of methods for identifying the critical effect.

It is widely agreed that the risk assessment of a specific substance at expected human exposure levels can be improved if extended use is made of the dose-response curves from studies in experimental animals for that substance, in order to better characterise and quantify potential risks (FOSIE 2002). The benchmark dose (BMD) approach provides a means of achieving this, see EPA 1995. The BMD approach estimates the dose that causes a low but measurable target organ effect (e.g. a 5% reduction in body or organ weight or a 10% increase in the incidence of kidney toxicity). By calculating the lower confidence limit of that estimated dose, the uncertainty and variability in the data is taken into account. While this lower confidence limit was originally termed the benchmark dose (Crump, 1984), it was later denoted and subsequently became known as the benchmark dose lower confidence limit (BMDL).

The BMD approach has also been applied to human data, derived from epidemiological studies. Additional points have to be considered when a BMDL is calculated using epidemiology data for risk assessment such as the uncertainties associated with the quantitative estimation of human exposure, and the effect of confounding factors.

Neither the former EC Scientific Committee on Food nor EFSA have systematically used the BMD approach so far. However, experimental data from which a BMD has been calculated are submitted from time to time. The BMDL has been proposed by the EFSA as the reference point for the estimation of margins of exposure for substances that are both genotoxic and carcinogenic, an approach that was used by the Joint FAO/WHO Expert Committee on Food Additives (JECFA) at its meeting in 2005 (JECFA, 2006a). There is a clear 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 proposed task is to assess the existing information on the utility of the BMD approach and to make recommendations on whether EFSA should use this approach and under which circumstances this is considered appropriate. Furthermore, guidance is needed for the selection of appropriate uncertainty factors when using a BMDL as a reference point.

TERMS OF REFERENCE AS PROVIDED BY EFSA

Following the suggestion of the Scientific Committee for a self-task on this subject, the European Food Safety Authority requests the Scientific Committee to:

- draft a document on advantages and limitations of the use of the benchmark dose approach as opposed to the traditionally used no-observed-adverse effect levels.
- prepare an opinion of the preferred approach to be used for risk assessments conducted by EFSA's Panels and Expert Groups, and
- in case the BMD approach would be considered the most appropriate approach for defined situations in regulatory risk assessments, to provide guidance on a harmonized introduction of this approach, as appropriate, in regulatory risk assessments carried out by EFSA's Panels and Expert Groups, including guidance for the selection of appropriate uncertainty factors when using BMDL as a starting point.

ACKNOWLEDGEMENTS

The European Food Safety Authority wishes to thank the members of the Working Group on Benchmark Dose for the preparation of this opinion: Ada Knaap (chair), Alan Boobis, Hans Bouwmeester, Paul Brantom, Esben Budtz- Jørgensen, Lutz Edler, Claudia Fruijtier-Pölloth (member of the working group until July 2008), Philippe Grandjean, Inge Mangelsdorf, Douglas McGregor, Andrew Renwick, Ivonne Rietjens, Salomon Sand, Josef Schlatter, Wout Slob.

The European Food Safety Authority would also like to thank Prof. David Bellinger (Department of Environmental Health – Harvard School of Public Health) for his contribution to section 5.7.

ASSESSMENT

1. Introduction

This document 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 opinion. Toxicity studies are conducted to both identify and characterize the potential adverse effects of a test material. Analysis of the data obtained in these studies is structured to identify a dose that can be used as a starting point for human health risk assessment. The dose used for this purpose, however derived, is referred to in this paper 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) is a RP that has been commonly used in risk assessment of non-genotoxic substances. The NOAEL has a long history of use in the regulatory process of human risk assessment, and is the usual RP for estimating 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. As the NOAEL is known to have some limitations (see following sections) alternative methods have been explored in recent years. The Benchmark Dose (BMD) approach arose as an alternative way of defining a RP. The present paper explores this method for deriving a RP and how it compares to the current convention of using a NOAEL for determining a RP.

Both the 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 for calculation of the margin of exposure for substances that are both genotoxic and carcinogenic. The US-EPA has used the BMD approach for many years for setting Reference Doses (RfDs) for non-genotoxic compounds (although the NOAEL is also used for this purpose, on a case by case basis) and to calculate the reference point and cancer slope factors for carcinogens. Outside the U.S. there has not been widespread use of the BMD approach for setting a RP for non-genotoxic substances, although in individual cases (e.g. World Health Organization (WHO), European Medicines Agency (EMA), EFSA) the BMD approach was part of the risk assessment.

In sections 1 to 3 of this 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 4 discusses the potential impact of using the BMD approach for hazard/risk characterisation and risk communication while Section 5 summarizes how the BMD approach may be applied in practice, and provides a number of recommendations regarding its application. In the Appendix, a practical guidance is given on how to operate the software and how to interpret the output generated by the software.

2. Hazard identification: selection of potential critical endpoints

Toxicity studies are designed to identify the adverse effects produced by a substance, and to define the dose-response relationships for the adverse effects detected. While in some cases human dose-response data are available, most risk assessments rely on 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, e.g. pair-wise statistical evaluations of the data. Alternatively dose-response modelling approaches (like the BMD-approach) can be used.

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 data³ 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 analyzed in more detail as described in the next sections.

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 calculate 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 effects which are large enough to be observable given the small sample sizes used in animal studies. In addition, doses are needed to provide information on the lower part of the 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.

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

3.1. The NOAEL approach

The NOAEL approach is applicable to all toxicological effect 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 treatment level 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 observation 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 significant effect at all dose levels, the lowest dose used in the study may be set 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 steepness of the dose-response curve close to the NOAEL. In situations where the dose-response relationship is not monotonic over the lower doses, it is possible that a dose group will be ignored for identification of the NOAEL, even when the response is statistically significantly different from that of the background response. Similarly, when the magnitude of the response is low, 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 LOAEL but the dose-response relationship is very steep, the NOAEL may be identified as the next dose down (i.e. 2 dose groups below the LOAEL). Such decisions require expert judgement and it is clear that 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).

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, derived from the estimated dose-response curve, associated with a specified change in response, the Benchmark Response (BMR), (see Section 3.5). 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 response is likely to be smaller than 5%

where the term “likely” is defined by the statistical confidence level, usually 95%-confidence.

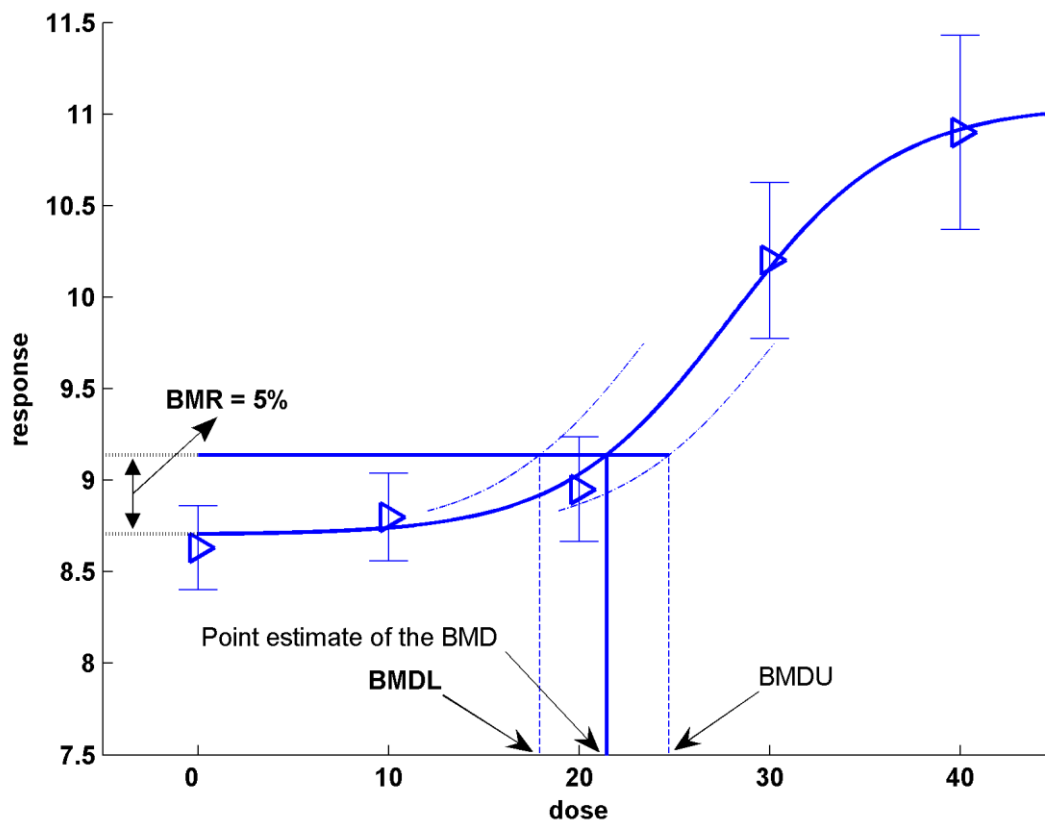


Figure 1: Key concepts for the BMD approach, illustrated by using hypothetical continuous data.

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 BMD (point estimate), 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.

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 ($8.7 + 0.05 \cdot 8.7 = 9.14$). Thus, the BMD₀₅, 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 BMDL₀₅ has a value of 18.

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

- Specification of a low but measurable response level, e.g. a 5% or 10% increase or decrease in response compared with the background response. This is called the BMR.
- Fitting a set of dose-response models, and calculation of the BMD and the BMDL for those models that describe the data according to statistical criteria, resulting in a range of BMDL values for each adverse effect/endpoint (see sections 5.3 & 5.4).
- Selection of a BMDL for each potentially critical endpoint, see section 5.5
- An overall study BMDL, i.e. the critical BMDL of the study, is obtained from the range of BMDL values for the different potentially critical endpoints (see section 5.5).

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.

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 control group, high limit of detection of quantitative 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 BMDL for that dataset will tend to be low. But the meaning of that 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.

⁴ In this opinion, a lower (or upper) 95%-confidence bound (one-sided) is equivalent to the lower (or upper) limit of a two-sided 90%-confidence interval.

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. Some guidance on this issue is given in section 5.5

The most well known BMD software is 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)

3.3. Interpretation and properties of the NOAEL and the BMD

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.

The NOAEL is not necessarily a “no effect” dose, although it is sometimes interpreted as such. However, as the animal studies discussed in section 3.5 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). 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) 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 response⁶. For an individual NOAEL the size of effect remaining statistically non-significant might be smaller, or greater than these values. This is illustrated below (Table 1), where for a number of recently evaluated substances the upper bound of effect at the NOAEL was calculated. Such additional information is, however, not reported in current applications of the NOAEL approach. In the BMD approach, the potential size of the effect (i.e. the benchmark response, BMR) is by definition known.

For a limited number of compounds, the Scientific Committee determined upper bounds for the effect size that are summarized in table 1. 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 5.4). Similarly, for this limited number of compounds, 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 5.4). In one of the examples, with a highly variable clinical-chemistry parameter, the upper bound of effect was as high as 260%.

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%.

Table 1: Illustrations of upper bounds⁷ of effect at NOAELs related to 10 compounds evaluated previously by JMPR or EFSA.

Compound (source +year)	Endpoint	Quantal data	Continuous data	References
		Upper bound extra risk (%) ⁸	Upper bound percent change (%) ⁹	
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 corpuscular 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

⁷ As calculated by the Scientific Committee.

⁸ Two-sided 90%-confidence interval for extra risk was calculated by the likelihood profile method

⁹ 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 the use of 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 the NOAELs represent 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: it is reflected by the confidence interval around the BMD. This is not normally done with the NOAEL approach. This is a difference that is particularly important in cases where the data are limited, or show much scatter. Although in such cases a NOAEL can often be derived, the uncertainty in the value obtained is usually not quantified.

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 data quality 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 BMDL has already accounted for the quality of the particular dataset (by being a lower bound estimate), so that data quality 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 for the BMD approach. As these guidelines are revised, e.g. within the OECD Test Guidelines Programme, the need to provide data better-suited for the BMD approach should be taken into account.

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 dose-response data.

Example 1: Continuous dose-response data

This example relates to relative kidney weights, which are continuous response data. In continuous data a numerical value is measured for each subject. 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 300 and 600 mg/kg are significantly different at $p < 0.05$, while the effect at the lowest dose (75 mg/kg) is non-significant. Based on the criterion of a statistically significant result, 75 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 7.4% increase in relative kidney weight (Table 2).

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

Dose	N	Geometric mean	ES (%)	Lower 95%-confidence bound (one sided) of ES (%)**	Upper 95%-confidence bound (one sided) of ES (%)**	t-statistic	p-value
0	5	7.16					
75	5	7.35	2.6	-2.0	7.4	0.99	0.17
300	5	7.72	7.8	3.0	12.9	2.88	0.006
600	4*	8.19	14.4	9.0	20.1	4.85	< 0.001

GM = geometric mean

ES = effect size (in percent change compared to response at dose zero)

* = one animal died at the top dose

** 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)

To illustrate the BMD approach for the same dataset, a dose-response model ($y = a \exp(b x)$) was fitted to the data, and a BMR representing a 5% increase in relative kidney weight was used (see Fig. 2). The output of this model results in a BMD_{05} of 221 mg/kg and a $BMDL_{05}$ (at $BMR = 5\%$) of 171 mg/kg (see legend of Fig. 2).

In this dataset the $BMDL_{05}$ is higher than the NOAEL (171 vs. 75 mg/kg). Nonetheless, it can be stated that the effect size at the $BMDL_{05}$ is smaller than 5% (with 95% confidence), while according to the pair-wise comparison (see Table 2) it can only be stated that the effect size at the NOAEL is smaller than 7.4% (again with 95% confidence). In other words, the confidence in the reference point provided by the BMDL is greater than that provided by the NOAEL. This illustrates that the BMD approach makes better use of the information in data by analysing the complete dataset, rather than making comparisons between single dose groups and the background response.

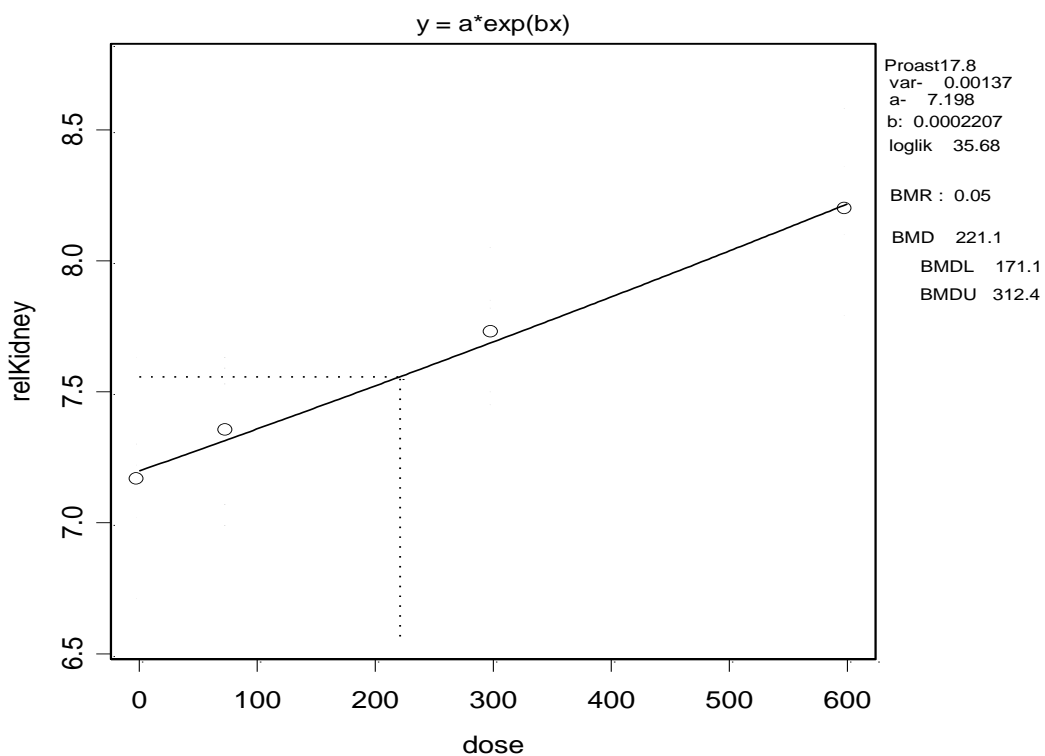


Figure 2: Relative kidney weights (mg/kg bw) in five individual animals per dose.

Larger circles represent group (geometric) means. The solid curve is the fitted dose-response model and the dashed lines indicate the BMD at a BMR of 5%. At the time of the preparation of this document, this feature was not implemented in BMDS.

BMDU = upper confidence bound for BMD.

Example 2: Quantal response data

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

Here, the mid-dose and low-dose incidences are not statistically significantly different from the background response. Hence, the middle dose of 450 mg/kg is the NOAEL for this endpoint in this study. In this case, a pair-wise comparison with the background response results in a very large upper 95%-confidence bound (one sided) for the effect size at the NOAEL: an extra risk¹⁰ value of around 47%.

Modelling the dose-response data (see Fig. 3) using a log-logistic model as an illustration results in a BMD₁₀ of 399, and a BMDL₁₀ of 171 mg/kg, considerably lower than the NOAEL.

The BMD methodology allows for the statement that the associated effect at the BMDL is not greater than 10% (with 95% confidence), which is considerably lower than the upper bound of effect of around 47% at the NOAEL, as calculated based on a pair-wise comparison of the background response and the NOAEL dose group. Again, this illustrates the fact that the BMD approach makes better use of the available information in the data.

¹⁰ 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.

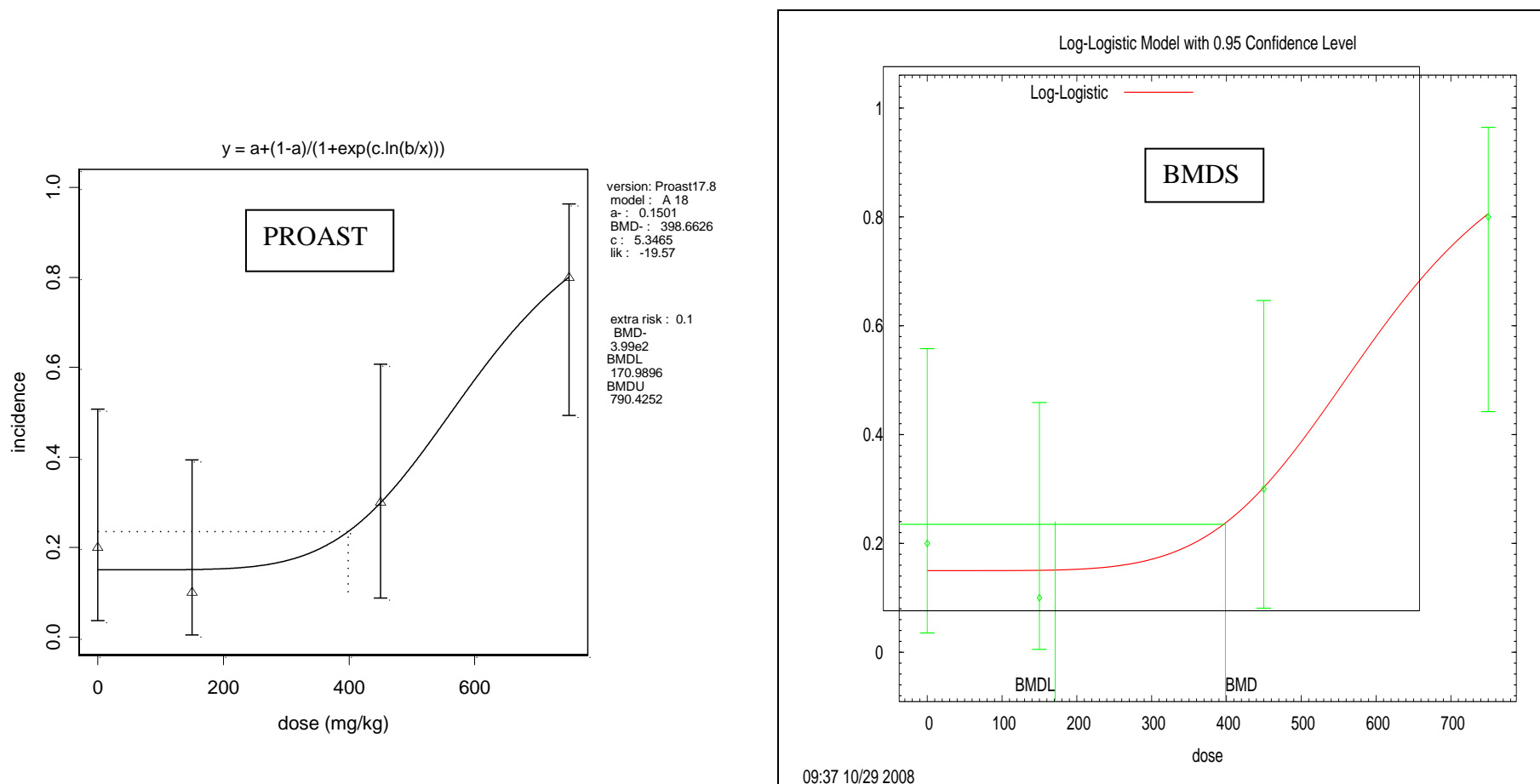


Figure 3: Analysis of quantal data as obtained by PROAST and BMDS software.

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% extra risk compared to the response at zero dose (according to the curve). Log-logistic model was fitted by PROAST and BMDS (see Table 3); the figures presented 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 the presence of confounders, 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 categorised into a limited number of dose groups. 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, suggesting 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 5.7).

The example further illustrates that the BMD approach may apply in situations without any controls: the background response level can be estimated by the fitted dose-response curve.

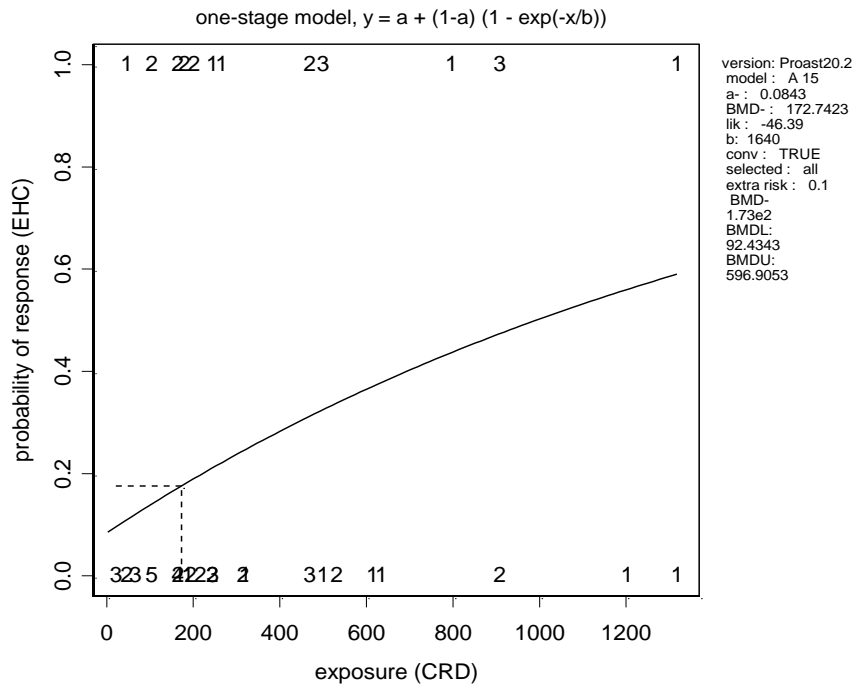


Figure 4: BMD analysis of human dose-response data with individual exposures.

Observed eye-hand coordination scores (0.0 = normal, 1.0 = abnormal) in individual workers (the plotted numbers indicate the number of persons at that exposure level) as a function of exposure. A dose-response model has been fitted to these data; the BMD_{10} (see dashed lines) was 173, and the $BMDL_{10}$ was 92. A BMR of 10% extra risk was used.

3.5. Selection of the benchmark response (BMR)

Derivation of a RP using the BMD approach requires the selection of a specified value for the benchmark response (BMR). Here, default values for the BMR are proposed with a brief justification for their choice.

Ideally, the BMR would reflect an effect size that is negligible or non-adverse. However, the practical constraint is that the BMR chosen should not be too small, to avoid having to estimate a BMD by extrapolation outside the range of observation, such that the BMDL would then depend heavily on the model used.

Initially, the BMD approach was mainly applied to quantal data, and BMR values of 1%, 5% or 10% extra risk have been proposed to be used for such data from animal studies (Crump, 1984; EPA, 1995). Use of a 5% BMR is practicable for some but not all toxicity datasets and 10% appears more appropriate for quantal data because the BMDL can become substantially dependent on the choice of dose-response model at lower BMRs (Sand et al., 2002). The BMDL₁₀ has been reported to be on average close to the NOAELs for lethality data (Fowles et al., 1999). Similarly, for quantal data in developmental toxicity, it was found that the BMDL was, on average closer to the average NOAEL for a BMR of 10% than for a BMR of 5% or 1%. But in this case the BMDL₁₀ was still on average 2-fold lower than the NOAEL (Allen et al. 1994). In human studies BMRs as low as 1% have been applied, which might be justified if the number of subjects is very large (e.g., a BMR of 1% was used for deriving BMD/BMDL for aflatoxins from human studies – EFSA 2007).

For continuous data, a 5% response level is often within the range of observation, and hence this BMR would provide estimates of BMD and BMDL that are not critically model dependent (Woutersen et al., 2001; Piersma et al. 2000; Sand et al. 2006). A re-analysis of a large number of National Toxicology Program (NTP) studies (Bokkers and Slob, 2007) showed that the BMDL₀₅ 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).

In conclusion the Scientific Committee proposes that a default BMR value of 10% be used for quantal data and 5% for continuous data from animal studies. As stated previously, the default BMR may be modified based on statistical or toxicological considerations.

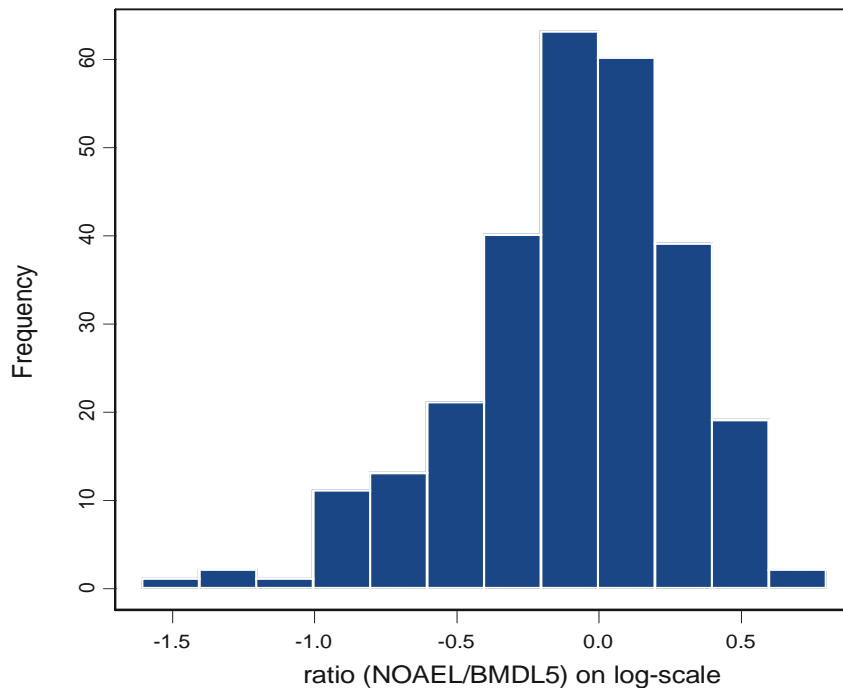


Figure 5: Histogram of 395 NOAEL/BMDL05 ratios (log₁₀-scale) for the same dose-response data in rat and mouse (NTP) studies.

The BMDL 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 BMDL₀₅. Data from Bokkers and Slob (2007).

4. Consequences for hazard/risk characterisation and risk communication

In the previous section the BMD approach has been introduced in the context of deriving a reference point (RP). The next step in hazard characterisation is the setting of margins of exposure (MOEs) for substances that are both genotoxic and carcinogenic or 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, for substances which are not both genotoxic and carcinogenic.

4.1. Setting health-based guidance values

In deriving an ADI/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 effect. As discussed (in Section 3.5), the default values of the BMR are such that the BMDL on average coincides with the

NOAEL. Further, it was shown in section 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 BMD/BMDL. Taking these considerations into account, an additional uncertainty factor, beyond those normally applied is not necessary. 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 or biologically significantly different from the response in the control group (LOAEL). In the NOAEL approach, the LOAEL is traditionally divided by an additional uncertainty factor. However in the BMD approach, it is often 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 dose-response data do not allow a BMR at the desired level (see section 3.5), then a BMR at a higher response can be selected and an additional uncertainty factor to the BMDL may be necessary.

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.

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 interpolation between applied doses. For the same reasons the BMD approach is also suitable for the derivation of 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 organophosphates¹¹). 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). The PPR Panel has published a recent Opinion in which the BMD, rather than BMDL, was preferred for determining relative potency estimates (EFSA, 2008).

¹¹ See http://www.epa.gov/pesticides/cumulative/pr-a-op/i_b-f.pdf

The BMD approach is suitable for analysing the effects of covariates on the dose-response, such as sex, exposure duration, or co-exposure to another chemical. For example, the factor sex can be included as a covariate in the dose-response analysis, and in this way it can be assessed if males and females show statistically significantly different dose-responses. The finding that one sex is more sensitive to the chemical than the other may be useful information in the risk assessment. In addition, when the dose-responses for the levels of males and females or other subpopulations are found to have a similar form, the dose-response data from both sexes can be combined. This effectively increases the overall sample size of the analysis, while in some cases it may happen that the total number of applied doses effectively increases. Likewise dose-response data from different studies with similar characteristics could be combined, with study as a covariate, again to make use of all data available (Murata et al., 2002).

While the use of covariates in the statistical analysis has so far been limited in the analysis of animal studies, it is common practice in the analysis of human data. The latter is further discussed in section 5.7

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). 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).

4.5. BMD 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 BMD 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 4.1, use of the BMDL in risk assessment does not change fundamentally 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 quality 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 4.1, the NOAEL and BMDL are expected to be similar on average. A review would certainly not be necessary in circumstances where large margins exist between the estimated daily intake (EDI) and the health based guidance value, e.g. ADI. For substances where the actual EDI 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, i.e. more or less conservative. Such information will help to determine whether the likely refinement provided by the BMD approach will result in a meaningful change in the risk assessment. The foregoing is relevant when the only objective of using the BMD approach is to improve confidence in the reference point.

In addition, when an health-based guidance value is based on the BMD approach and 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.

5. Guidance to applying the BMD approach

This section provides an overview of how to derive a BMD and BMDL from dose-response data and recommendations are given on particular choices to be made. 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
2. Specification of the BMR
3. Selection of candidate dose-response model(s)

4. Identification of acceptable models
5. Estimating the BMD, and establishing the BMDL as the RP

These steps are discussed below.

5.1. Specification of type of dose-response data

Endpoints not showing dose response relationships are normally not used for deriving a BMD. The decision to disregard endpoints can be done by visual inspection of the data or more formally by including them in the BMD analysis in which case the decision is made on a statistically non-significant trend in step 3.

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 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. For quantal data the number of affected individuals and the sample size are needed for each dose group

5.2. Specification of BMR

For quantal data the BMR is defined in terms of an increase in the incidence of the lesion / response scored, compared with the background response. The common way of doing this is either by additional risk or extra risk, see section 3.4. For quantal data, the recommended default value is a BMR of 10% extra risk.

For continuous data the BMR could be defined in various ways. The way recommended here is to define it as a percent change in the average magnitude of the response variable as compared to the predicted background response. The recommended default value is a BMR of 5% (e.g. a 5% decrease in red blood cells).

For some continuous data (e.g. cognitive scores) the BMR may best be defined in a similar way as for quantal data, i.e. in terms of an increase by, say, 5% of the number of subjects with an abnormal response. This type of BMR is related to the so-called hybrid approach (Crump 2002).

5.3. Selection of candidate dose-response models

Different models which fit the data equally well can result in different BMDs and BMDLs, reflecting model uncertainty. To take this aspect of uncertainty into account, various models need to be fitted to the same dataset. This section provides a discussion of candidate models, in particular those that are currently available in the BMD software packages (e.g. BMDS, PROAST). This software provides adequate and flexible sets of models for analysing the range of datasets. If other software is used, it is recommended to apply the same set of candidate models.

Table 3 summarizes these models. Some of them are listed as part of a family of nested models, i.e., a simpler model (having few parameters) can be extended to a more complex model (having more parameters) such that it includes the simpler model. Table 3 illustrates the nested structure for the exponential family of models (see also below).

The advantage of using a family of nested models is that it can be determined which model has the appropriate number of parameters: no more than needed, but also not less than that. For instance, when the dose-response is nearly linear, a two-parameter model would suffice, while for a sigmoidal dose-response a three-parameter model at least would be needed (one for the location where the response reaches half its maximum, one for the magnitude of the maximum response, and one for the steepness of the curve between minimum and maximum response). The statistically optimal model can be found by starting with a simple model (with few parameters), and then checking whether adding a parameter to the model results in a significant improvement of the fit. If so, the model with the additional parameter is accepted. This process may be continued until no improvement in the fit can be found, resulting in the selection of a model with the optimal number of parameters. The decision whether or not a fit has been improved by the addition of a parameter can be based on significance testing, using the likelihood ratio test (see section 5.4).

Continuous data

For continuous data both the exponential family of models and the Hill family are recommended due to the following favourable properties:

- they are always monotonic,
- they are suitable for data that level off to a maximum response, or show a threshold-like curvature (see Fig. 6),
- they have been shown to be flexible enough to describe virtually all dose-response data as encountered in practice,
- they allow for incorporating covariates in a toxicologically meaningful way (see below).

It should be noted that the Hill family is not currently available as a nested series in the BMDS software.

Table 3 gives each of the members of both the exponential and Hill families. The difference between the BMDL for the optimal exponential model and the BMDL for the optimal Hill model provides an indication of the degree of model uncertainty in the particular dataset (see, for an illustration, Fig. 7 and table 5).

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, models with an additive background parameter would be needed. Alternatively, the body weight gains could be expressed as ratios (percent changes) rather than differences, if the individual body weight data are available.

The linear model often gives a sufficient description of epidemiological data and may therefore be used for deriving a BMD.

Polynomial models of a higher degree, although common in many statistical analyses, are not recommended for dose-response data, as it has several unrealistic model properties (e.g. they can result in negative values, which normally is biologically impossible, or they can take a non-monotonic shape even when the data do not support this).

Quantal data

Table 3 lists seven models that are recommended to be used for quantal data. Only the linearized multi-stage model (LMS) forms a nested family of models. Therefore, only one multi-stage model should be selected from this family and this member used further in the BMD analysis.

Modelling constraints for analysis of continuous and quantal data

To avoid the models having undesirable properties, certain constraints are imposed on the model parameters. For instance, since continuous responses are usually not negative, 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). Further, to avoid curves that have an infinite slope at dose zero (which might be considered biologically implausible for most endpoints) the “shape” parameter in some of the models is often constrained as follows:

(continuous data) M3 and M5 in exponential and Hill family:	$d > 1$
(quantal data) Log-logistic, Weibull, Gamma ¹² :	$c > 1$

¹² It should be noted that such a constraint is not possible for the log-probit model, as it has the property to always have zero slope at dose zero, even though this may not always be apparent from the plot.

Table 3: Recommended models for use in the BMD approach⁴⁾.

Model	Number of model parameters	Model expression response (y) as function of dose (x)	Constraints
<u>Continuous data</u>			
Exponential family			
Model 1 ¹⁾	1	$y = a$	$a > 0$
Model 2	2	$y = a \exp(bx)$	$a > 0$
Model 3	3	$y = a \exp(bx^d)$	$a > 0, d > 1$
Model 4	3	$y = a [c - (c-1)\exp(-bx)]$	$a > 0, b > 0, c > 0$
Model 5	4	$y = a [c - (c-1)\exp(-bx^d)]$	$a > 0, b > 0, c > 0, d > 1$
Hill family			
Model 2	2	$y = a [1 - x/(b+x)]$	$a > 0$
Model 3	3	$y = a [1 - x^d/(b^d+x^d)]$	$a > 0, d > 1$
Model 4	3	$y = a [1 + (c-1)x/(b+x)]$	$a > 0, b > 0, c > 0$
Model 5	4	$y = a [1 + (c-1)x^d/(b^d+x^d)]$	$a > 0, b > 0, c > 0, d > 1$
<u>Quantal data²⁾</u>			
Logistic	2	$y = 1 / (1 + \exp(-a - bx))$	$b > 0$
Probit	2	$y = \text{CumNorm}(a + bx)$	$b > 0$
Log-logistic	3	$y = a + (1-a) / (1 + \exp(-\log(x/b) / c))$	$0 \leq a \leq 1, b > 0, c > 1$
Log-probit	3	$y = a + (1-a) \text{CumNorm}(\log(x/b) / c)$	$0 \leq a \leq 1, b > 0, c > 0$
Weibull	3	$y = a + (1-a) \exp(-(x/b)^c)$	$0 \leq a \leq 1, b > 0, c > 1$
Gamma	3	$y = a + (1-a) \text{CumGam}(bx^c)$	$0 \leq a \leq 1, b > 0, c > 1$
Linearized multistage (LMS) family ³⁾			
one-stage	2	$y = a + (1-a) \exp(-bx)$	$a > 0, b > 0$
Two-stage	3	$y = a + (1-a) \exp(-bx - cx^2)$	$a > 0, b > 0, c > 0$
Three-stage	4	$y = a + (1-a) \exp(-bx - cx^2 - dx^3)$	$a > 0, b > 0, c > 0, d > 0$

a, b, c, d : unknown parameters that are estimated by fitting the model to the data.

CumNorm: cumulative (standard) normal distribution function.

CumGam: cumulative Gamma distribution function

¹⁾ Model 1 can be regarded as a model that is nested within any dose-response model: it reflects the situation of no dose-response (= horizontal line).

²⁾ For the constraints given here, the models result in increasing dose-response curves

³⁾ The one-stage model is identical to the quantal linear model as implemented in BMDS; note that in BMDS, this model is called “multistage” and the number of stages has to be defined by setting the degree of the polynomial in this model, e.g. 2 for a two-stage model.

⁴⁾ In epidemiology, additional models, e.g. $y=a+bx$, are also used

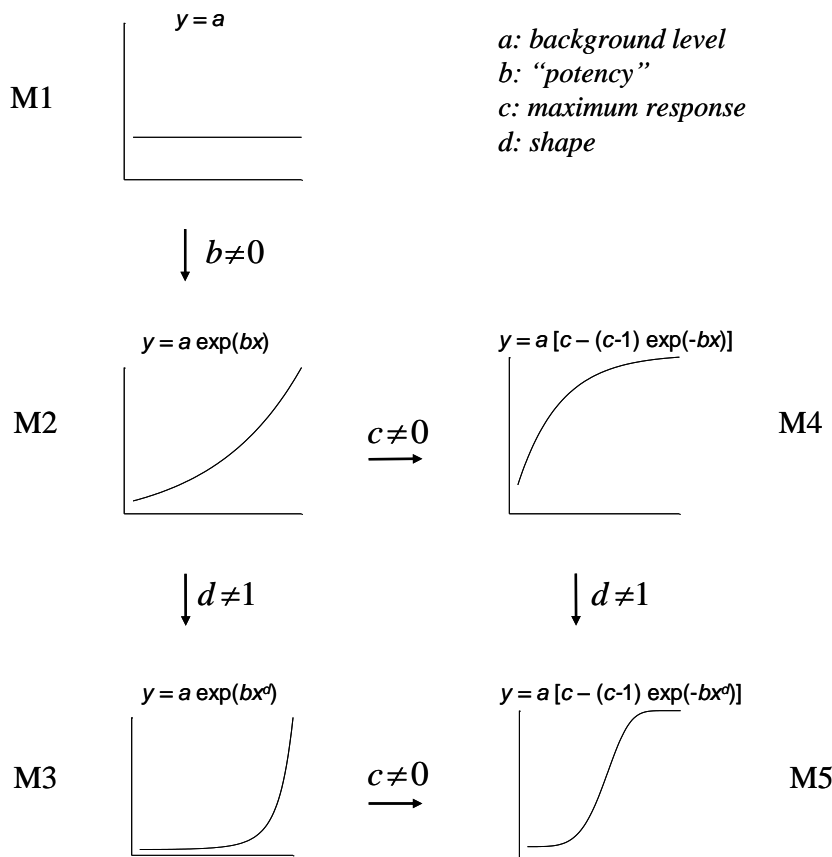


Figure 6: The exponential model family presented as a nested family of models.

Note that models M3 and M4 are not nested to each other. When a model results in a significantly better fit than its ‘predecessor’, the latter is rejected. These models can also be used for decreasing responses. Note that rejecting model M1 implies a statistically significant dose-related effect.

5.4. Fitting and accepting models

The available BMD software 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 and is achieved by maximizing the log-likelihood.

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 find those models with an acceptable fit.

The acceptance of a fitted model is based on two principles. The first principle is that from a nested family of models only one member is accepted, by comparing the log-likelihoods of the various members in that family, using the likelihood ratio test. As mentioned before, when a member with fewer parameters does not show a significantly poorer fit, then this member will be preferred.

The second principle is that any fitted model should provide reasonable description of the dose-response data, according to a goodness-of-fit test with a P value greater than 0.05. There are several types of goodness-of-fit tests. The likelihood ratio test is the recommended choice here. If this is not available in the software package used, other tests, such as the Pearson chi-square test should be considered. In the likelihood ratio test, the log-likelihood value associated with a fitted model is compared with, and tested against, the log-likelihood value associated with the so-called “full model”. The full model simply consists of the observed (mean) responses at each applied dose. Hence, the number of parameters equals the number of dose groups. If a model’s fit is not significantly worse than that of the full model, then the model may be accepted.

The likelihood ratio test may be used to test if additional parameters in nested models result in a significant improvement of the fit. This test is based on the fact that minus twice the difference of the log-likelihood values associated with two models follows a Chi-square distribution with number of degrees of freedom equal to the difference in number of parameters between the two models. Table 4 provides the critical difference between two log-likelihoods, given a particular increase in the number of parameters (degrees of freedom). See Slob (2002) for more details.

Table 4: Critical difference in log-likelihood values in comparing two nested models.

df	Critical difference in log-likelihoods	df	Critical difference in log-likelihoods
1	1.92	11	9.84
2	3.00	12	10.51
3	3.91	13	11.18
4	4.74	14	11.84
5	5.54	15	12.50
6	6.30	16	13.15
7	7.03	17	13.79
8	7.75	18	14.43
9	8.46	19	15.07
10	9.15	20	15.71

When the difference exceeds the critical value, the fit of the model with more parameters is better at a 5% -significance level. The number of degrees of freedom (df) is equal to the difference in the number of estimated parameters between both models.

It is important to keep in mind that there may be cases where none of the models pass the goodness of fit test, while visual inspection of the data appears to show that some models adequately describe the observed dose-response. In those cases it might be that the assumptions of the goodness-of-fit test are violated (e.g., deviances from the assumed distribution, not all of the observations are independent, non-random errors) and the decision to accept a particular model needs to acknowledge the high level of uncertainty in the assessment.

5.5. Estimating the BMD, and establishing the BMDL as the Reference Point

The BMD analysis should be performed for potentially critical endpoint(s) measured in a study:

- For each potentially critical endpoint, apply the set of models (table 3), and for each of the accepted models, calculate the BMD, together with its confidence interval. The lower bound of this confidence interval is the BMDL. The default confidence level is 95% (one-sided). This procedure results in a range of BMDL values, reflecting the differences between the models used.
- Determine the lowest BMDL from this range, to be the overall BMDL for each potentially relevant critical endpoint, i.e. the most conservative value in each case.
- Determine the lowest BMDL from all these endpoints, which will then be the overall BMDL for that study.

Selecting the lowest BMDL amongst those for the same endpoint tends to be conservative, but it is recommended to use this until more advanced methods, such as “model averaging” (Wheeler and Bailer 2007, 2008) have been fully developed and validated.

In this way, BMDLs can be determined for various endpoints. If the critical endpoint was not established before the BMD analysis, it may then be determined based on the outcomes from the BMD analysis in a way analogous to that in the NOAEL approach, i.e. choose the endpoint resulting in the lowest BMDL, provided that the effect is relevant to human health.

When the experimental data provide sufficient information on the dose-response relationship, the various models that fit the data will have similar shapes and will yield a narrow range of BMDL values. Such BMDL values should provide a secure basis for risk assessment purposes and can be used to define a RP for the establishment of a health-based guidance value or for the calculation of a MOE (see Section 4).

In some circumstances, the dose response relationship may not be well defined. For instance, there may be large gaps between consecutive response levels. In such datasets the various models (see Table 3) that fit the data (according to the statistical criteria discussed above) may assume different shapes, and consequently the ranges of BMDL values obtained may be wide. These BMDL values would not provide a secure basis for establishing a health-based guidance value. Criteria to judge the adequacy of the dose-response data on the basis of the range of BMDL values obtained have so far not been established. The Scientific Committee proposes that, as a general rule, dose-response data should not result in a range of BMDL values from different accepted models that substantially exceeds one order of magnitude. When this value is exceeded, several options are available and should be considered on a case-by-case basis, e.g. increasing the BMR, re-evaluating the set of models, or model averaging.

5.6. Illustrations

Continuous data

A BMD analysis for a single dose-response data set is illustrated in Fig. 7 and Table 5. Note that in this example dataset, the number of dose levels (6) is larger than in typical toxicity studies. Both the exponential and the Hill model families were applied to the dose-response data set. The table shows the log-likelihood values achieved by each member of both model families, as well as by the full model. In this analysis, M3 is selected as the “optimal” model within both families of models (in most cases the selected models from different families have the same number of parameters). These two models result in BMDLs of 326 and 316 mg/kg respectively, and this difference reflects model uncertainty (which is small in this case, as a result of the relatively large number of dose groups). Thus, the BMDL for this endpoint would be 316 mg/kg, being the lowest BMDL.

The selection of the optimal model within a nested family can be done by some BMD software, but the user can do it by him/herself. As an illustration consider the log-likelihood values¹³ in the third column of Table 5. Keeping in mind which models are nested within others (see Fig. 6) the following procedure may be followed. The log-likelihood associated with M2 is more than 8 units higher than M1, while it has one more parameter. According to the likelihood ratio test, adding a parameter to a model results in a significantly ($p=0.05$) better fit when the log-likelihood has increased by 1.92 or more. Hence, M2 is clearly better than M1. Based on the same criterion, M3 is better than M2, while M4 is not better than M2 (the additional parameter c was estimated to be zero, in which case M4 is equivalent to M2). Further, M5 is not better than M3. Hence, M3 is the ‘optimal’ model from this nested family. Finally, M3 cannot be significantly improved and passes the goodness of fit test: the required increase in the log-likelihood value would be 4.74 for an additional 4 parameters.

¹³ It should be noted that log-likelihood values have no absolute meaning (the value depends on the number of data points, for instance). They only have relative meaning, when comparing their values for different models applied to the same dataset.

Table 5: Results of BMD analysis using two families of models.

Model ¹⁾	Number of model parameters	Log-likelihood		BMD ₀₅ * (mg/kg bw)		BMDL ₀₅ * (mg/kg bw)	
		Exponential	Hill	Exponential	Hill	Exponential	Hill
M1 (null model)	1	20.85					
└-- M2	2	28.97	28.61				
└-- M3	3	33.35 **	33.39 **	488	479	326	316
└-- M4	3	28.97	28.61				
└-- M5	4	33.56	33.56				
Full model	7	35.95					

* At BMR of 5%.

** Optimal model within this family according to log-likelihood criterion. It also passes the goodness-of-fit test, as the full model does not result in a significantly better fit.

¹⁾ See Figure 7 and Table 3 for an illustration of models M1 to M5, and their nested structure

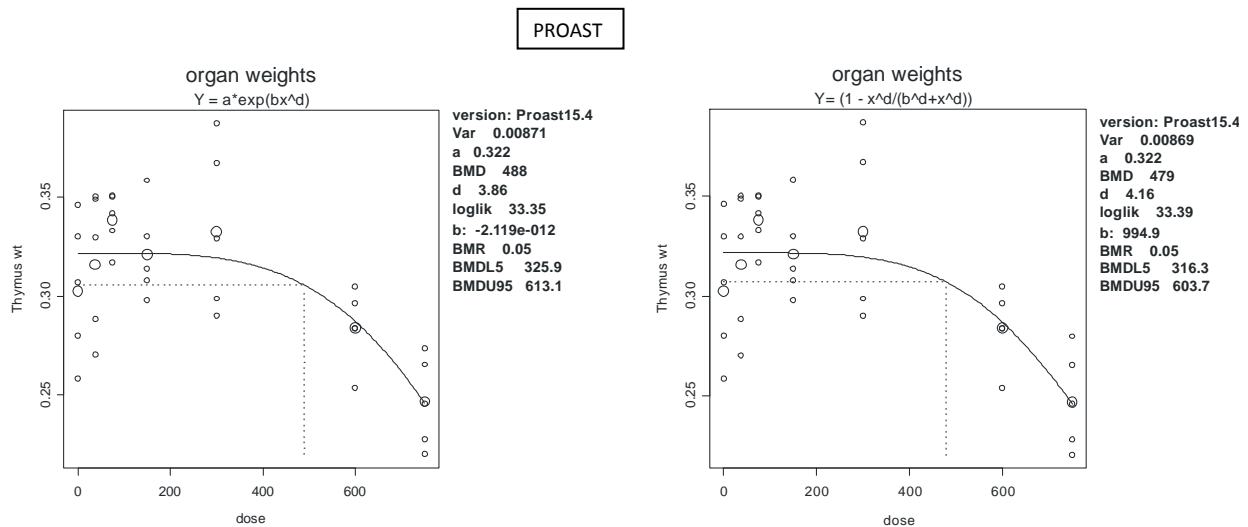


Figure 7: Fitted curves for M3 in the exponential model family (left panel) and for M3 in the Hill model family (right panel). See Table 3.

Quantal data

Table 6 and Fig. 8 show the results for various models fitted to a quantal dataset. The determination of accepted models is based on the three criteria as discussed above:

- All models are highly significantly better than the no-response model: the log-likelihood values are around 50 units higher, where an increase of only 1.92 units (for models with 2 parameters) or 3.00 units (for the models with 3 parameters) would have been sufficient (see table 4).
- For the (nested) LMS model, the two-stage model is significantly better than the one-stage model, since the difference in log-likelihoods is larger than 1.92. Therefore, the two-stage model represents the LMS model in this case.
- Only three of the seven models passed the goodness of fit test: the other models show log-likelihoods that are significantly lower than the log-likelihood of the full model.

The associated BMDLs for the three accepted models range from 1.93 to 2.45 mg/kg, and this small range indicates that these dose-response data are suitable for deriving a BMDL₁₀.

Table 6: Results of BMD analysis of quantal data shown in Fig. 8.

model	number of model parameters	log-likelihood	accepted	BMD ₁₀	BMDL ₁₀
null model	1	-143.01	--		
LMS: one-stage	2	-56.8	no		
LMS: two-stage	3	-53.71	no		
logistic	2	-55.63	no		
probit	2	-60.40	no		
Weibull	3	-52.33	no		
log-logistic	3	-49.35	yes	3.28	2.45
log-probit	3	-49.56	yes	3.16	2.42
gamma	3	-50.96	yes	2.88	1.93
full model	4	-49.33	--		

BMR: 0.10

Constraint: no

P-value Goodness of Fit: <0.05 leads to model rejection. If the model passes the goodness-of-fit test, see section 5.4

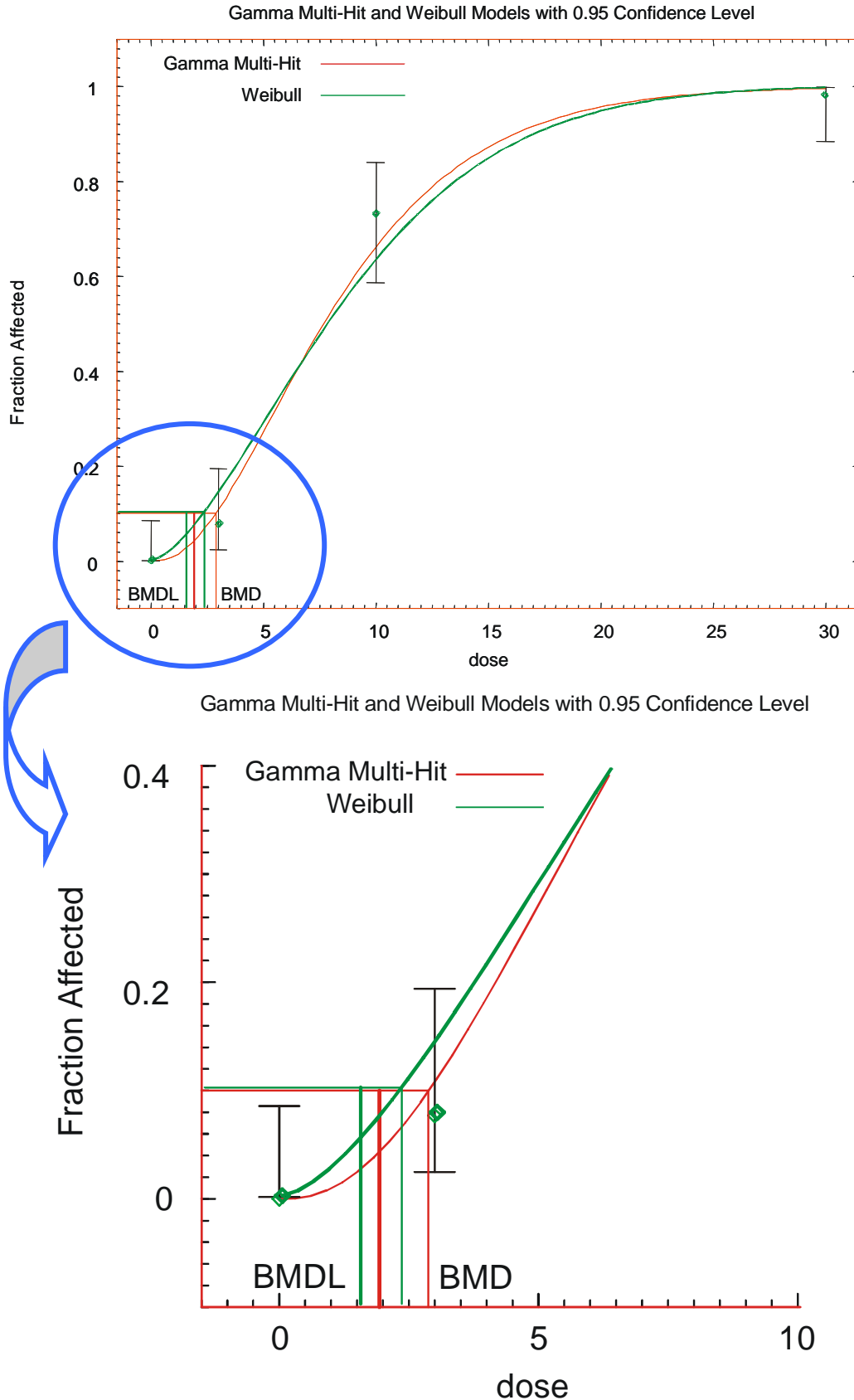


Figure 8: Two examples of the quantal model, Gamma and Weibull fitted to a tumour incidence dose-response dataset, using the BMDS software

The Gamma model was accepted ($P=0.071$), while the Weibull model was rejected ($P=0.014$).

5.7. Specific issues of human dose-response data

Dose-response data from observational epidemiological studies may differ from typical animal toxicity data in several respects. The main differences relevant to BMD calculations are briefly discussed below.

Exposure data often do not fall into a small number of well defined dosage groups. To apply the NOAEL approach, individual exposure data need to be categorised with a consequent loss of information. As demonstrated in Example 3, the BMD approach can be readily used with individual exposure data.

Unlike most experimental studies, observational studies may not include an unexposed control group, because all individuals may be exposed to some extent, e.g. an atmospheric pollutant, a food contaminant. In this case, the BMD approach still applies, since fitting a dose-response curve does not necessarily require observations at zero exposure. However the response at zero exposure would then need to be estimated by low-dose extrapolation. Hence the BMD derived from epidemiological data can be strongly model-dependent (Budtz-Jørgensen et al., 2001).

It should also be noted that the estimation of human exposure is often imprecise, and ignoring the imprecision may lead to a biased assessment of the dose-response relationship. Statistical methods have been developed to tackle this issue and obtain less biased estimates of the dose-response relationship and of the BMD value (Budtz-Jørgensen et al., 2004; Budtz-Jørgensen, 2007).

Response variables in human studies are often subject to confounding factors that may interfere with the dose-response of interest. Failure to take a confounding factor into account may result in either underestimation or overestimation of the BMD. This is commonly addressed by including the confounding factors as covariates, e.g. in a multiple regression model. Effect modification may present an additional issue that needs to be taken into account, e.g. when a greater vulnerability occurs in elderly subjects. The BMDL should then reflect the response in the most vulnerable subpopulation. Effect modification can be modelled by including the relevant covariates. Adjustment for confounding or effect modification is not possible in BMDS, and only partly in the PROAST software.

Despite these methodological problems, the BMD analysis has been used with observational epidemiological studies in a number of cases (e.g. EFSA, 2009b)

5.8. 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). 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 sizes¹⁴ should be given for each dose level. For human data similar information should be provided along with details on confounding factors..
- B. The value of the BMR chosen, and the rationale for it.
- C. The software used, including version number
- D. Settings and assumptions of the model fitting procedure
- E. A table presenting the models used and the log-likelihoods, information on fitting and accepting models, and the BMDs and the BMDLs for the accepted models.
- F. At least one plot of a fitted model to the data for the critical endpoint(s), including the one for the lowest BMDL
- G. Conclusion regarding the BMDL of this particular study.

The reporting of a BMD analysis is illustrated below for a quantal and for a continuous dataset. In both cases one single endpoint is analysed, which is assumed to represent the critical effect. The reader is referred to the appendix for a brief guidance on how to use the BMDS and the PROAST software.

Whilst efforts have been made in this opinion to provide guidance on the use of BMD software, users should be aware that such software is still evolving and new versions are released frequently. Hence, the version of the software available at the time of use may not be the same as that referred to here. Users are advised to consult the manual for the respective version of the software being used. Regardless of the version of the software used, the reporting structure should remain the same.

Example 1: Continuous data.

This example was run using PROAST software. Note that if using the BMDS software, some of the default settings need to be modified to those recommended in this document (see appendix).

¹⁴ 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.

This example relates to a 2-year study with deoxynivalenol in mice. A dose-related decrease of body weight was observed. This endpoint is assumed to be the critical effect. 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.

A. The data

dose	body weight, group mean	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 17.3

D. Additional assumptions: None

E. Tables of BMD results

Body weight (males)

Model	N° of parameters (variance excluded)	Log-likelihood		BMD ₀₅ * (mg/kg)		BMDL ₀₅ * (mg/kg)	
		Exponential	Hill	Exponential	Hill	Exponential	Hill
M1 (no response)	1	119.03					
└-- M2**	2	197.91	197.43	0.25	0.22	0.23	0.20
└-- M3	3						
└-- M4	3						
└-- M5	4						
Full model	7	198.34					

* At BMR of 5%.

** Optimal model within this family according to log-likelihood criterion. It also passes the goodness-of-fit test, as the full model does not result in a significantly better fit. Consequently, there was no need to fit the higher models. M3, M4 and M5

F. Figures

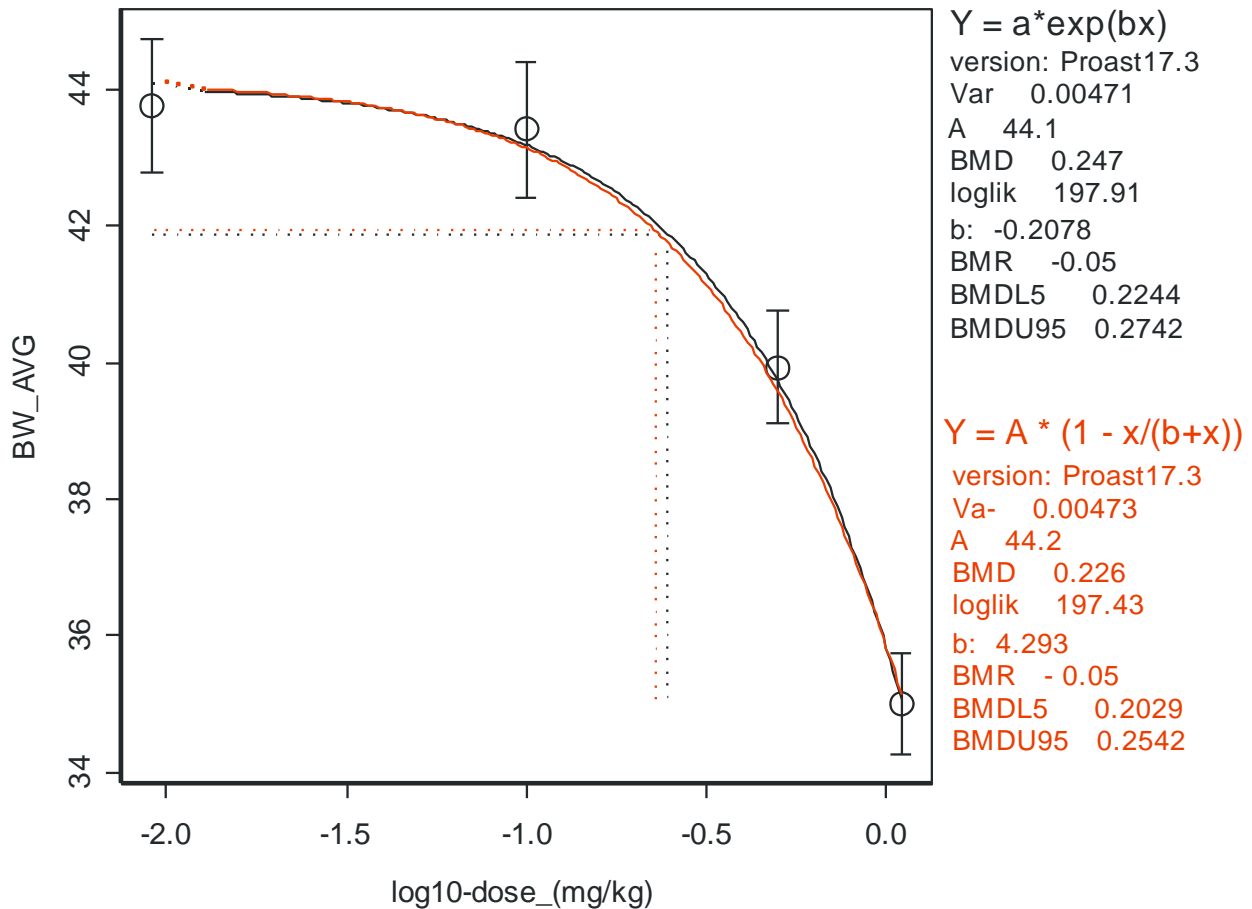


Figure 9: Fitted curves for M2 from the exponential model family (black) and for M2 from the Hill model family (red).

On these figures the doses have been plotted on a log-scale for better readability; the controls have been plotted at an arbitrary low dose.

G. Conclusion

For both the exponential and the Hill family of models, model 2 was selected. The associated $BMDL_{05}$ values were 0.23 and 0.20 mg/kg. The $BMDL_{05}$ for this study is 0.20 mg/kg.

Example 2: quantal data

This example relates to a 2-year study in rats, where three doses of a compound were administered to the animals. Dose-related changes in thyroid epithelial cell vacuolization were found, and these data were used here for a BMD analysis.

A. The data

Dose (mg/kg bw/day)	N° of animals showing 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

B. BMR: Default value (extra risk = 0.10)

C. Software used: BMDS version 2.0.

D. Additional assumptions: None

E. Tables of BMD results

Model used	N° of parameters	log-likelihood	accepted	BMD ₁₀	BMDL ₁₀
Null model	1	-137.19			
Gamma	3	-93.49	no		
Logistic	2	-97.23	no		
Log-logistic	2	-91.91	yes *	3.23	1.90
Log-probit	3	-91.86	yes *	3.31	2.34
LMS: one-stage ¹⁾	2	-94.10	no		
LMS: two-stage	3	-94.10	no		
Probit	2	-97.54	no		
Weibull	3	-93.77	no		
full model	4	-90.02			

* not significantly worse than full model (and significantly better than null model)

¹⁾ The one-stage model is identical to the quantal linear model as implemented in BMDS; note that in BMDS, this model is called “multistage” and the number of stages has to be defined by setting the degree of the polynomial in this model, e.g. 2 for a two-stage model.

F. Figures

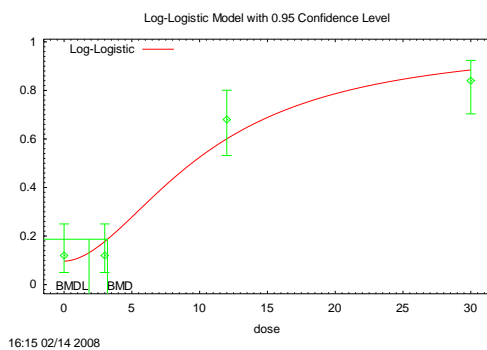


Figure 10: Plot of log-logistic model fitted to the data. The log-probit model resulted in a similar plot.

G. Conclusion

Two of the models were accepted, with an associated $BMDL_{10}$ range of 1.9 to 2.3 mg/kg bw. The $BMDL_{10}$ for this study is: 1.9 mg/kg bw.

CONCLUSIONS AND RECOMMENDATIONS

The Scientific Committee concludes that the BMD approach is a scientifically more advanced method compared with 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 and variability in the dose-response data. Using the BMD approach results in a more consistent RP, as a consequence of the specified BMR.

Whilst software is currently available for BMD analysis, the Scientific Committee notes that this is evolving and significant developments are anticipated in the medium term, e.g. model averaging, analysis of continuous data.

The Scientific Committee also considers that when toxicity test guidelines, e.g. OECD guidelines, are revised, specific aspects of the BMD approach are taken into account.

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. Where refinement of previous risk assessments is considered necessary, application of the BMD approach would be of particular value.

The BMD approach is applicable to all chemicals in food, independently of their category or origin, e.g. pesticides, additives or contaminants. Examples where the BMD approach adds particular value include situations where the identification of a NOAEL is uncertain, where a Reference Point is needed for the Margin of Exposure calculation for substances that are genotoxic and carcinogenic, and for dose response assessment of epidemiological data. In the short term, the Scientific Panels and EFSA Units are strongly encouraged to adopt the BMD approach for situations such as those described above.

In the longer term, the Scientific Committee anticipates that the BMD approach will be used as the method of choice for the determination of the Reference Points for deriving health-based guidance values and margins of exposure. To that end, recognising that there are practical considerations regarding its introduction and wider use in EFSA, and recognising that its application requires a level of expert judgement and modelling expertise, the Scientific Committee proposes that training in dose-response modelling and the use of the software be offered to experts in the scientific Panels and EFSA Units. The Scientific Committee would then like to review the implementation, experience and acceptability of the BMD approach in EFSA's work in two years time.

REFERENCES

- Allen, B.C., Kavlock, R.J., Kimmel, C.A., and Faustman, E.M. (1994). Dose-response assessment for developmental toxicity. II. Comparison of generic Benchmark dose estimates with No Observed Adverse Effects Levels. *Fundam. Appl. Toxicol.* 23:487-495.
- Baird, S. J. S., Cohen, J. T., Graham, J.D., Shylakter, A. I., & Evans, J. S. (1996). Noncancer risk assessment: A probabilistic alternative to current practice. *Human and Ecological Risk Assessment*, 2, 79-102.
- Bokkers, B.G.H. and Slob W. (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the Benchmark dose approach. *Crit. Rev. Toxicol.* 37: 353-377.
- Boobis, A.R, Cohen, S.M, Dellarco, V., McGregor, D. Meek, M.E., Vickers, C, Willcocks, D and Farland W. (2006) IPCS Framework for analysing the relevance of a cancer mode of action for humans. *Crit.Rev.Toxicol*,36, 781-792.
- 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* (in press).
- Crump, K.S. (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. *Crit. Rev Toxicol* 32, 133-153.
- EFSA (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* 282, 1-31
- EFSA (2007). Opinion of the scientific panel on contaminants in the food chain [CONTAM] related to the potential increase of consumer health risk by a possible increase of the existing maximum levels for aflatoxins in almonds, hazelnuts and pistachios and derived products. *The EFSA Journal* 446, 1-127
- EFSA (2008). Opinion of the Scientific Panel on Plant Protection products and their Residues to evaluate the suitability of existing methodologies and, if appropriate, the identification of new approaches to assess cumulative and synergistic risks from pesticides to human health with a view to set MRLs for those pesticides in the frame of Regulation (EC) 396/2005. *The EFSA Journal* 704, 1-84
- EFSA (2009a). Safety evaluation of ractopamine. Scientific Opinion of the Panel on Additives and Products or Substances used in Animal Feed. *The EFSA Journal* 1041, 1-52.
- EFSA (2009b). Cadmium in food. Scientific Opinion of the Panel on Contaminants in the Food Chain. *The EFSA Journal* 980, 1-139.
- EPA (1995). The use of the benchmark dose approach in health risk assessment. EPA/630/R-94/007. Risk Assessment Forum, Washington DC.

- FOSIE (2002). Food Safety in Europe: Risk assessment of chemicals in food and diet. *Food and Chemical Toxicology*, 40 (2-3).
- Fowles JR, Alexeeff GV, Dodge D. (1999). The use of benchmark dose methodology with acute inhalation lethality data. *Regul Toxicol Pharmacol* **29**, 262-278.
- Fryer, M., Collins, C. D., Ferrier, H., Colvile, R. N., & Nieuwenhuijsen, M. J. (2006). Human exposure modeling for chemical risk assessment: A review of current approaches and research and policy implications. *Environmental Science & Policy*, 9, 261–274.
- Gibney, M. J., & van der Voet, H. (2003). Introduction to the Monte Carlo project and the approach to the validation of probabilistic models of dietary exposure to selected food chemicals. *Food Additives and Contaminants*, 20(suppl. 1), S1–S7.
- Joint FAO/WHO Expert committee on food Additives – JECFA (2006a), Sixty-fourth meeting, WHO/IPCS Safety evaluation of certain contaminants in food. WHO Food Additives Series 55.
- Joint FAO/WHO Expert committee on food Additives – JECFA (2006b). WHO Technical Report Series 939. Evaluation of certain veterinary drug residues in food. 66th report of the Joint FAO/WHO Expert Committee on food additives.
- Kavlock RJ, Allen BC, Faustman EM, Kimmel CA. (1995). Dose-response assessments for developmental toxicity IV. Benchmark doses for fetal weight changes. *Fundam Appl Toxicol* **26**, 211-222.
- Murata K, Budtz-Jorgensen E, Grandjean P. (2002). Benchmark dose calculations for methylmercury-associated delays on evoked potential latencies in children. *Risk Anal*; 22: 465-74.
- Piersma, A.H., Verhoef, A., te Biesebeek, J.D., Pieters, M.N. and Slob, W. (2000). Developmental toxicity of butyl benzyl phthalate in the rat using a multiple dose study design. *Reprod. Toxicol.*, 14 (5), 417-425.
- Pieters, M.N., Bakker, M. and Slob, W. (2004). Reduced intake of deoxynivalenol in The Netherlands: a risk assessment update. *Toxicol Lett.*, 153, 145-153.
- Sand S, Falk Filipsson A, Victorin K. (2002). Evaluation of the benchmark dose method for dichotomous data: model dependence and model selection. *Regul Toxicol Pharmacol* **36**, 184-197.
- Sand S, von Rosen D, Victorin K, Falk Filipsson A. (2006). Identification of a critical dose level for risk assessment: developments in benchmark dose analysis of continuous endpoints. *Toxicol Sci* **90**, 241-251.
- Slob W, and Pieters MN. (1998). A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. *Risk Anal* **18**, 787-798.
- Slob, W. (2002). Dose-response modelling of continuous endpoints. *Toxicol. Sci.*, 66 (2), 298-312.
- Swartout J.C., Price P.S., Dourson M.L., Carlson-Lynch H.L., Keenan R.E. (1998). A Probabilistic Framework for the Reference Dose (Probabilistic RfD). *Risk Anal* 18: 271-282

- Tressou, J., Leblanc, J. C., Feinberg, M., & Bertail, P. (2004). Statistical methodology to evaluate food exposure to a contaminant and influence of sanitary limits: Application to Ochratoxin A. *Regulatory Toxicology and Pharmacology*, 40, 252–263.
- Wheeler, M.W. and Bailer, A.J. (2007). Properties of Model-Averaged BMDLs: A Study of Model Averaging in Dichotomous Response Risk Estimation. *Risk Analysis*. 27, 659 – 670.
- Wheeler, M.W. and Bailer, A.J. (2008). Model averaging software for dichotomous dose response risk estimation. *Journal of Statistical Software*. 26(5), 1 – 15.
- WHO (1987): Principles for the Safety Assessment of Food Additives and Contaminants in Food. Environmental Health Criteria 70. WHO/IPCS
- Woutersen, R.A., Jonker, D., Stevenson, H., J.D. te Biesebeek, 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

APPENDIX: SOFTWARE FOR BMD ANALYSIS

[The appendix consisting of practical guidance on how to use available software, the Scientific Committee only took note of this section of the document, which was therefore not subject to formal adoption.]

Dose-response analysis and modelling can be performed in many statistical software packages. However, for the non-expert, it is recommended that software that has been specifically developed for the purpose of BMD analysis is used. Such software packages are up to date with the various aspects of the BMD concept, including the statistical methodology, suggested dose-response models, and procedures for the calculation of the BMD and BMDL. Also, the terminology in the field is reflected in these software packages. The most well known BMD software is 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).

The output results generated by the BMDS and PROAST software will be illustrated and discussed in the sections 1 and 2 of this appendix respectively. In principle, the two software programs should provide the same outcome when applied to the same data set provided that the same dose-response model and other assumptions/settings are used. Negligible differences may potentially result due to differences in the numerical algorithms employed. The BMDS and PROAST software are both applicable to quantal as well as continuous dose-response data, but they differ in a number of aspects, e.g. in terms of the user interface, and to some extent regarding the models and procedures available. The BMDS and PROAST software will be discussed in sections 1 and 2 below. In section 3, the major differences between the two software packages are summarized.

1 The benchmark dose software, BMDS

The BMDS is easily accessible. It can be downloaded free of charge from the U.S. EPA's website. It is a Windows-based package. Version 2.0 of the BMDS is considered in the discussion below. The user interface of BMDS 2.0 differs from older versions and more advanced features are included in BMDS 2.0. However, the basic elements generated by BMDS 2.0 (the type of plots and numerical output) discussed in sections 1.3 and 1.4 are identical to those in older versions of the software. It should be pointed out, however, that future versions of the BMDS may potentially differ somewhat from the illustrations provided below, which are based on version 2.0.

In section 1.1, the different dose-response models in the BMDS are described, and in section 1.2 the different definitions available for the BMR are presented. In sections 1.3 and 1.4, the output results generated by the BMDS 2.0 are illustrated and discussed for quantal and continuous data, respectively. The case where a specific model is fitted to an individual data set is considered. More advanced options also exist in the BMDS; e.g. several models may be fitted at the same time to the same data set. These advanced features are illustrated in section 1.5. Together with the illustrations in sections 1.3, 1.4, and 1.5, guidance on how to operate the BMDS is also given. For more detailed information on how the software is operated the help manual of the BMDS can be consulted.

1.1 Dose-response models in the BMDS (version 2.0)

Several different models are available in the BMDS for both continuous and quantal (referred to as Dichotomous in the software) data. The different models that are available under each category are summarized in Table 1 below.

Table A1. The different dose-response model available in the BMDS.

Continuous	Quantal (Dichotomous)	Dichotomous_Alternative ¹⁾	Dichotomous_Nested ¹⁾
Hill	Gamma	Gamma with background dose	NLogistic
Exponential	Weibull	Weibull with background dose	NCTR
Power ²⁾	Quantal linear ²⁾	Multistage with background dose	Rai and Van Ryzin
Linear ²⁾	Multistage ⁴⁾	Multistage-cancer with background dose	
Polynomial ²⁾	Multistage-cancer ³⁾	log-Probit with background dose	
	log-Probit	Probit with background response	
	log-Logistic	Logistic with background response	
	Probit Logistic	Dichotomous-Hill	

¹⁾ These models are not discussed in the opinion

²⁾ Not recommended in this opinion

³⁾ Same as Multistage model but includes a slope factor for linear extrapolation

⁴⁾ Called LMS in the opinion

Continuous models

For continuous data, five types of models are available. The Hill and Exponential models are sigmoidal and can produce s-shaped dose-response curves. They are recommended in this opinion. The Exponential model in BMDS is the family of exponential models 2, 3, 4, and 5 that are also available in PROAST (see section 2). At the date of the writing of this report only a *beta* version of the exponential family models exists in the BMDS and no plotting feature has been added for these models yet. Some restrictions can be made with respect to the continuous models:

-With the Hill model, there is an option to restrict the slope parameter, $\eta \geq 1$. This option is usually activated.

-With the Exponential model, there is an option to restrict “power” to ≥ 1 . This option is similar to that for the Hill model, and this option is usually activated.

For continuous data a variance model needs to be selected in addition to the model for the mean response. In the BMDS, the choice between a constant and a non-constant variance model can be made. The non-constant model corresponds to a power function of the mean response. Thus, the variance depends on the mean response for which a model should also be specified (e.g. the Hill model), and it has two additional parameters. The mean response and the variances are estimated together.

-If the option “rho = 0” is activated in the software the constant model will be selected and if this option is not activated the non-constant model will be used.

Quantal models

For quantal data several dose-response models are available. There are nine “traditional” quantal models (Dichotomous), eight alternative models (Dichotomous_Alterative), and three models for studies which include data relating to litter (Dichotomous_Nested). The default pool of models recommended in this opinion is the Weibull, gamma, multistage, log-logistic, logistic, log-probit, and probit models. The Weibull, gamma, log-logistic, and log-probit models have three parameters, including a parameter that explicitly describes the background response. Certain restrictions can be made for this group of models:

-With the Weibull and gamma models there is an option to restrict “power” to ≥ 1 . This option is usually activated.

-With log-logistic and log-probit models there is an option to restrict “slope” to ≥ 1 . This option is similar to that for the Weibull and gamma models, and is activated by default. However for the log-probit model, the slope should not be restricted.

The multistage models also have a parameter which describes the background response. The total number of parameters in these models depends on the degree of the polynomial in the exponential terms of the models. This degree needs to be specified by the user. Certain restrictions can be made for these models:

-With the Multistage models there is an option to restrict “Betas” to ≥ 0 . This option should be activated, as it prevents the curve from being non-monotonic.

The Probit and Logistic models have two parameters. In contrast to the other models they do not include a parameter that explicitly describes the background response.

1.2 Benchmark response (BMR) definitions in the BMDS

For quantal data, two types of BMR definitions are available; additional and extra risk. This opinion recommends an extra risk of 10 % as the default BMR for quantal data. An extra risk of 10 % is also the default BMR in the BMDS software.

For continuous data several BMR definition are available in the BMDS, which are summarized below:

- Std. Dev:** This implies a change in response relative to the estimated standard deviation in the control group. Std Dev = 1 is the default BMR in the BMDS software for continuous data.
- Rel. Dev:** This implies a % change in response relative to the estimated mean response in the control group. In this opinion, Rel Dev = 0.05 is recommended as the default for continuous data.
- Abs. Dev:** This implies an absolute change in response from the estimated mean response in the control group
- Point:** This implies that the continuous response value corresponding to the BMD (and BMDL) is directly specified.
- Extra:** This implies a percent change in response from the estimated difference between the maximum and minimum response value. This option is only available for the Hill and Exponential models which are sigmoidal and approach some limiting response value as the dose approaches infinity. This BMR definition is effectively similar to the extra risk definition for quantal data.

1.3 The BMDS and quantal data

The following section discusses and illustrates the BMDS for quantal data. First, a short guide is provided on how to proceed when using the BMDS 2.0 in the case of analysing a quantal data set with a particular dose-response model:

1. After starting the BMDS software, go to File-New Dataset and select the type of data to be used: Select “Dichotomous” (i.e. quantal data) and the model that should be fitted (i.e. gamma, logistic, log-logistic, probit, log-probit, Weibull, or multistage). Insert the data column-wise; the doses, the number of animals per group, and the incidence, or % positive, in each group.

2. Save the data under File-Save Dataset As

3. Click proceed, a new screen appears:

Under “Column Assignment”, indicate the column number (or column name), in the data set, containing the doses, the number of animals, and the incidences (or % positive), respectively.

Under “Other Assignments”, activate the box for “BMD calculation”, and “BMDL Curve Calc.” if requested. The model specific restrictions discussed in section 1.1 can also be activated here (e.g. power ≥ 1). The suggested default values can be used for the Confidence Level (= 0.95), the BMR (= 0.10), and the Risk Type (= Extra).

4. Run the calculation

5. The BMDS will display a plot that shows the data and the fitted model graphically. The BMDS also presents a numerical output file in which the numerical values for the BMD, the BMDL, and details on statistics of the model fit can be found.

6. For further details and advanced options see section 1.5 and the BMDS user manual

In the example presented below the Weibull model has been fitted to a hypothetical quantal data set. In the analysis, the power term was restricted to ≥ 1 , the BMR was specified to an extra risk of 0.1 (i.e. 10%), and the default confidence level of 0.95 was used for the BMDL calculation. The resulting BMDS plot is shown in Figure A1, and the numerical output is given in Table 2. In the numerical output, information is first given regarding the model selected, the data used as basis, and the iteration settings applied in the model fitting process. Further down in Table 2, statistical information of the model fit and performance is given, and the BMD results from the analysis are finally presented. The more important parts of the numerical output; i.e. *Parameter Estimates*, *Analysis of Deviance Table*, *Goodness of Fit*, and *BMD computation*, is discussed in detail below (Tables A3-A6).

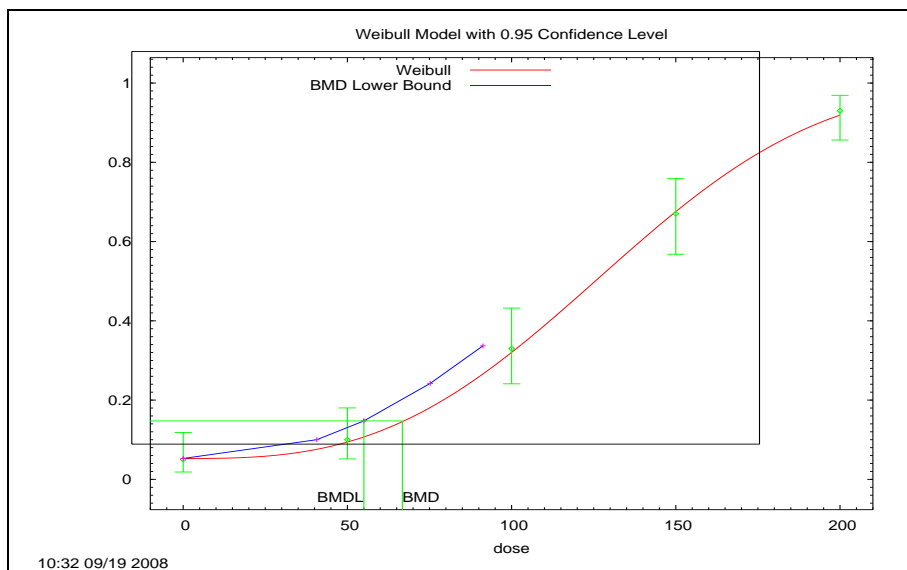


Figure A1. Plot generated by the BMDS.

The Weibull model has been fitted to a hypothetical quantal data set. The BMR has been set to an extra risk of 0.1. A one-sided 0.95 (i.e. 95%) confidence level was used for the BMDL calculation. The Weibull model has 3 parameters: “background”, “slope”, and “power”.

Table A2. BMDs numerical output for quantal data.

```

=====
Weibull Model using Weibull Model (Version: 2.12; Date: 05/16/2008)
Input Data File: C:\USEPA\BMDs2\Data\lWeil-2T.o.(d)
Gnuplot Plotting File: C:\USEPA\BMDs2\Data\lWeil-2T.o.plt
                               Fri Nov 07 10:32:57 2008
=====

BMDs Model Run
~~~~~

The form of the probability function is:

P[response] = background + (1-background)*[1-EXP(-slope*dose^power)]

Dependent variable = EFFECT2
Independent variable = DOSE
Power parameter is restricted as power >=1

Total number of observations = 5
Total number of records with missing values = 0
Maximum number of iterations = 250
Relative Function Convergence has been set to: 1e-008
Parameter Convergence has been set to: 1e-008

Default Initial (and Specified) Parameter Values
Background = 0.0544554
Slope = 4.35787e-007
Power = 2.94054

Asymptotic Correlation Matrix of Parameter Estimates

Background      Slope      Power
Background      1          -0.48      0.47
Slope           -0.48      1          -1
Power           0.47      -1          1

Parameter Estimates

Variable      Estimate      Std. Err.      95.0% Wald Confidence Interval
Background    0.0527242     0.020575      Lower Conf. Limit  Upper Conf. Limit
Slope        5.51437e-007  9.89168e-007  -1.3873e-006      2.49017e-006
Power        2.89478      0.353106      2.2027           3.58685

Analysis of Deviance Table

Model      Log(likelihood)  # Param's  Deviance  Test d.f.  P-value
Full model      -204.559         5          0.1936    2          0.9077
Fitted model    -204.656         3          269.849   4          <.0001
Reduced model    -339.484         1

AIC:          415.312

```

Goodness of Fit					
Dose	Est._Prob.	Expected	Observed	Size	Scaled Residual
0.0000	0.0527	5.272	5.000	100	-0.122
50.0000	0.0950	9.501	10.000	100	0.170
100.0000	0.3255	32.553	33.000	100	0.095
150.0000	0.6842	68.420	67.000	100	-0.305
200.0000	0.9243	92.425	93.000	100	0.217
Chi^2 = 0.19		d.f. = 2	P-value = 0.9078		
Benchmark Dose Computation					
Specified effect =	0.1				
Risk Type =	Extra risk				
Confidence level =	0.95				
BMD =	66.7396				
BMDL =	55.0227				

Table A3. Parameter Estimates.

Parameter Estimates				
Variable	Estimate	Std. Err.	95.0% Wald Confidence Interval	
			Lower Conf. Limit	Upper Conf. Limit
Background	0.0527242	0.020575	0.0123979	0.0930504
Slope	5.51437e-007	9.89168e-007	-1.3873e-006	2.49017e-006
Power	2.89478	0.353106	2.2027	3.58685

In the *Parameter Estimates* section in the numerical output (Table A3), the parameter estimates associated with the fitted dose-response model are given. Standard errors (Std. Err.) and 95 % confidence intervals for the estimated parameter are also given. Table A3 should not be mixed up with the *Default Initial (and Specified) Parameter Values* given earlier in the numerical output (Table A2), which are the start values for the estimation process. The quantal models in the BMDs have in general 2 or 3 parameters that are given descriptive names in the numerical output; e.g. background, slope, and power (Table A3), and relate to the parameters in the “probability function” shown in Table A2.

Table A4. Analysis of Deviance

Analysis of Deviance Table					
Model	Log(likelihood)	# Param's	Deviance	Test d.f.	P-value
Full model	-204.559	5			
Fitted model	-204.656	3	0.1936	2	0.9077
Reduced model	-339.484	1	269.849	4	<.0001
AIC:	415.312				

In the *Analysis of Deviance Table* (Table A4), three log-likelihood values are given. The log-likelihood can be seen as quantitative measure of how close the model is to the data points; the higher the log-likelihood (i.e. less negative in this example), the better the fit. The log-likelihood values for three different “models” are shown here, which can appear somewhat confusing since only one model, the Weibull model (“Fitted” model), was specified by the

user. However, the “Full” and “Reduced” models are theoretical reference cases that are used for comparison. These different models are described below and illustrated in Figure A2:

Full model: A theoretical model which fits the data points perfectly, also called the “saturated model”. It has one parameter for each dose group but it does not correspond to some mathematical equation, as the Weibull model. Rather it corresponds to the observations (in this case the observed incidences) and no statement is given of how this model looks between the data points. The log-likelihood associated with this model denotes the maximum that can be obtained for the particular dataset studied; if the Weibull model fitted the data perfectly it would have had the same log-likelihood.

Reduced model: A theoretical model which has only one parameter; it just corresponds to a horizontal line, the average incidence. This is a model that describes a “no dose-response relationship”, and is also called the “null model” in the opinion.

Fitted model: This is the user specified model, the Weibull model in this example.

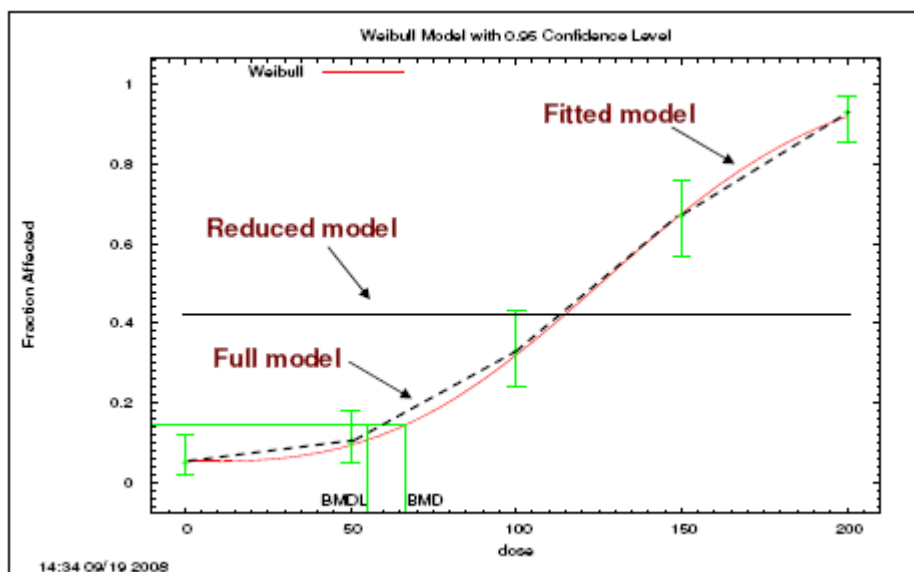


Figure A2. Illustration of the “Fitted model” (the user specified Weibull model); the “Full model” (a model that fits the data perfectly); and the “Reduced model” (Null model).

The “Full” and “Reduced” models are theoretical reference cases. The “Full” and “Reduced” models do not appear in the BMDS plot, but have been added here for illustrative purposes.

According to a likelihood ratio test, the difference between the log-likelihoods associated with two models is used to formally investigate whether or not the models are significantly different, see table 4 in the opinion for the critical values (at $P=0.05$). Table A4 reports the deviance, which is twice the difference between the log-likelihoods from two models. The deviance is used for calculating the P-values as reported.

The p -value for the Fitted vs. Full model is > 0.05 (p -value = 0.9077, Table A4); the Weibull model is accepted since it is not significantly different from the “Full model” (the data points).

The p -value for the Reduced vs. Full model is ≤ 0.05 (p -value < 0.0001 , Table A4); there appears to be a dose-response trend since the horizontal line is significantly different from the “Full model” (the data points). For the particular example considered, the conclusion of a dose-response trend can, of course, be drawn directly by just looking at the data (Figure A2).

The Akaike’s Information Criterion (AIC) is also given in the *Analysis of Deviance Table* (Table A4). The AIC is a model selection criterion that can be used to compare models that belong to different classes or families, but is not recommended in this opinion.

Table A5. Goodness of Fit

Goodness of Fit					
Dose	Est._Prob.	Expected	Observed	Size	Scaled Residual
0.0000	0.0527	5.272	5.000	100	-0.122
50.0000	0.0950	9.501	10.000	100	0.170
100.0000	0.3255	32.553	33.000	100	0.095
150.0000	0.6842	68.420	67.000	100	-0.305
200.0000	0.9243	92.425	93.000	100	0.217
Chi ² = 0.19		d.f. = 2	P-value = 0.9078		

In *Goodness of Fit* (Table A5), the data set under evaluation is presented as well as more detailed information of the model fit at each dose level (Est._Prob.). Furthest to the right, the “Scaled Residuals” are given. The U.S. EPA suggests that a scaled residual should not be larger than 2 or smaller than -2 for any dose group, which is not supported in this opinion. An overall p -value is also given in Table A5 based on the “Chi-square” statistic which is obtained by squaring the scaled residuals and adding them together. The Chi-square and its corresponding p -value (which is derived from a χ^2 distribution for the appropriate d.f.) should be interpreted similar to the Deviance and its corresponding p -value for the Full vs. Fitted model (in Table A4). Both values represent a measure of the overall fit. However, they will slightly differ because the test statistic is calculated differently; the equation for the Deviance is different from that for the Chi-square. This opinion recommends using the p -value based on the likelihood ratio test (0.9077 as shown in Table 4).

Table A6. Benchmark dose computation

Benchmark Dose Computation	
Specified effect =	0.1
Risk Type =	Extra risk
Confidence level =	0.95
BMD =	66.7396
BMDL =	55.0227

Finally, in Table A6 the selected value of the BMR (0.1, i.e. 10%), BMR type (extra risk), and confidence level (0.95) are given together with the resulting BMD and its lower bound, BMDL. In quoting these values, only two significant figures should be quoted, i.e. BMDL = 55.

1.4 The BMDS and continuous data

The following section discusses and illustrates the BMDS for continuous data. First, a short guide is provided on how to proceed when using the BMDS 2.0, in the case of analysing an individual continuous data set with a certain dose-response model:

1. After starting the BMDS software, go to File-New Dataset and select the type of data to be used: Select “Continuous” and the model that should be fitted (i.e. Hill, polynomial, linear, exponential, power). Insert the data column-wise; the doses, the no. of animals per group, and the standard deviation and the mean response in each group. Alternatively, individual data can be entered; the doses and responses for all individual subjects.

2. Save the data under File-Save Dataset As

3. Click proceed, a new screen appears:

Under “Column Assignment”, indicate the column number (or column name), in the data set, containing the doses, the number of animals, the standard deviations, and the mean responses, respectively.

Under “Other Assignments”, activate the box “BMD calculation”, and “BMDL Curve Calc.” if requested. The model specific restrictions discussed in section 1.1 can also be activated here (e.g. $\eta \geq 1$). If “rho” = 0 is activated, a constant variance models is selected (see section 1.1). The default values suggested in the opinion should be used: Confidence Level (= 0.95), BMR (= 0.05), and Risk Type (= Rel. Dev.).

4. Run the calculation

5. The BMDS will display a plot that graphically shows the data and the fitted model. The BMDS also present a numerical output file in which the numerical values for the BMD, the BMDL, and details on statistics of the model fit can be found.

6. For further details and advanced options see section 1.5 and the BMDS user manual

In the example presented below the Hill model has been fitted to a hypothetical continuous data set. In the analysis, the slope parameter η was restricted to ≥ 1 , and a constant variance model was selected, “rho = 0”. The BMR was specified to 5% change in response compared with the estimated mean response in the control group (Rel. Dev = 0.05), and the default confidence level of 0.95 was used for the BMDL calculation. The version used in this example does not allow a log-normal transformation of the response data and assumes a normal distribution. The resulting BMDS plot is shown in Figure A3, and the numerical output is given in Table A7. In the numerical output, information is first given regarding the model selected, the data used in the modelling as basis, and the iteration settings applied in the model fitting process. Further down in Table A7, statistical information of the model fit and performance is given, and the BMD results from the analysis are finally presented at the end. The more important parts of numerical output; i.e. *Parameter Estimates, Table of Data and Estimated Values of Interest, Likelihoods of Interest, Tests of Interest, and BMD computation* will be discussed in detail below (Tables A8-A12).

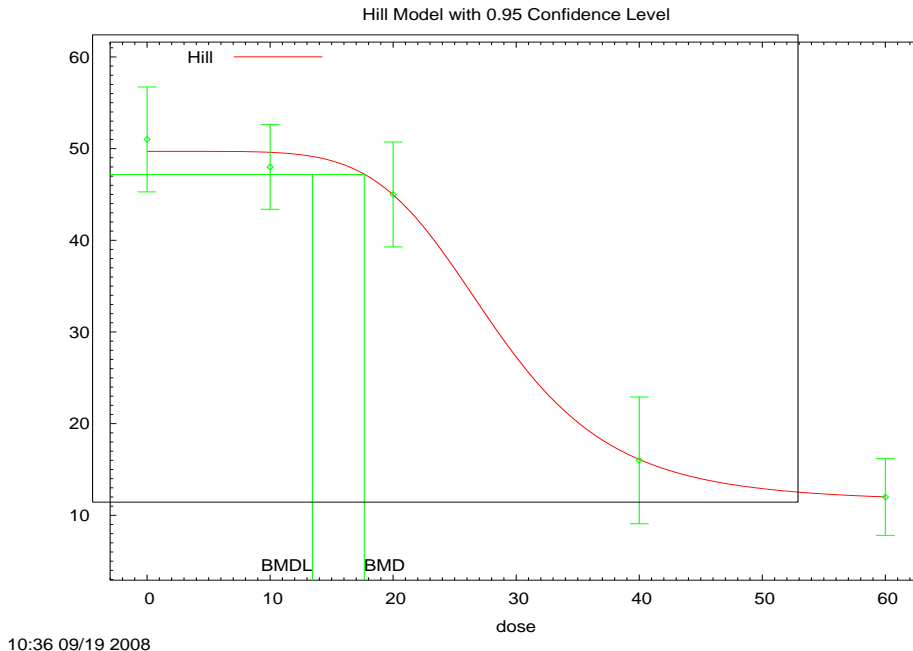


Figure A3. Plot generated by the BMDS.

The Hill model has been fitted to a hypothetical continuous data set. The BMR has been set in the options to a 5 %, i.e. =0.05 change in response relative to the estimated mean response in the control group. A 95% confidence level was used for the BMDL calculation. The Hill model has 4 parameters; “intercept”, “v”, “k” and “n”.

Table A7. BMDs numerical output for continuous data.

```

=====
Hill Model. (Version: 2.14; Date: 06/26/2008)
Input Data File: C:\USEPA\BMDs2\Data\8HilT.ds1..(d)
Gnuplot Plotting File: C:\USEPA\BMDs2\Data\8HilT.ds1..plt
                               Fri Nov 07 10:37:33 2008
=====
BMDs Model Run
~~~~~

The form of the response function is:

Y[dose] = intercept + v*dose^n/(k^n + dose^n)

Dependent variable = Mean
Independent variable = DOSE
rho is set to 0
Power parameter restricted to be greater than 1
A constant variance model is fit

Total number of dose groups = 5
Total number of records with missing values = 0
Maximum number of iterations = 250
Relative Function Convergence has been set to: 1e-008
Parameter Convergence has been set to: 1e-008

Default Initial Parameter Values
alpha = 55.1951
rho = 0 Specified
intercept = 51
v = -39
n = 4.58229
k = 29.3103

Asymptotic Correlation Matrix of Parameter Estimates

( *** The model parameter(s) -rho
have been estimated at a boundary point, or have been specified by the user,
and do not appear in the correlation matrix )

alpha      intercept      v      n      k
alpha      1      1e-008      -1.1e-007      -1.1e-007      1.3e-007
intercept  1e-008      1      -0.52      -0.34      -0.3
v          -1.1e-007      -0.52      1      0.72      -0.44
n          -1.1e-007      -0.34      0.72      1      -0.36
k          1.3e-007      -0.3      -0.44      -0.36      1

Parameter Estimates

Variable      Estimate      Std. Err.      95.0% Wald Confidence Interval
Lower Conf. Limit      Upper Conf. Limit
alpha      50.0614      10.4385      29.6023      70.5206
intercept  49.6614      1.66787      46.3924      52.9303
v          -38.2158      3.63855      -45.3472      -31.0843
n          5.69111      1.66023      2.43711      8.9451
k          28.2041      2.95884      22.4048      34.0033

```

Table of Data and Estimated Values of Interest

Dose	N	Obs Mean	Est Mean	Obs Std Dev	Est Std Dev	Scaled Res.
0	10	51	49.7	8	7.08	0.598
10	9	48	49.6	6	7.08	-0.66
20	10	45	44.9	8	7.08	0.0325
40	9	16	16	9	7.08	-0.02
60	8	12	12	5	7.08	0.0163

Model Descriptions for likelihoods calculated

Model A1: $Y_{ij} = \mu(i) + e(ij)$
 $\text{Var}\{e(ij)\} = \sigma^2$

Model A2: $Y_{ij} = \mu(i) + e(ij)$
 $\text{Var}\{e(ij)\} = \sigma(i)^2$

Model A3: $Y_{ij} = \mu(i) + e(ij)$
 $\text{Var}\{e(ij)\} = \sigma^2$
 Model A3 uses any fixed variance parameters that were specified by the user

Model R: $Y_i = \mu + e(i)$
 $\text{Var}\{e(i)\} = \sigma^2$

Likelihoods of Interest

Model	Log(likelihood)	# Param's	AIC
A1	-112.603521	6	237.207041
A2	-110.717413	10	241.434825
A3	-112.603521	6	237.207041
fitted	-113.004768	5	236.009537
R	-155.818533	2	315.637067

Explanation of Tests

- Test 1: Do responses and/or variances differ among Dose levels? (A2 vs. R)
 - Test 2: Are Variances Homogeneous? (A1 vs A2)
 - Test 3: Are variances adequately modeled? (A2 vs. A3)
 - Test 4: Does the Model for the Mean Fit? (A3 vs. fitted)
- (Note: When $\rho=0$ the results of Test 3 and Test 2 will be the same.)

Tests of Interest

Test	-2*log(Likelihood Ratio)	Test df	p-value
Test 1	90.2022	8	<.0001
Test 2	3.77222	4	0.4377
Test 3	3.77222	4	0.4377
Test 4	0.802495	1	0.3703

The p-value for Test 1 is less than .05. There appears to be a difference between response and/or variances among the dose levels. It seems appropriate to model the data.

The p-value for Test 2 is greater than .1. A homogeneous variance model appears to be appropriate here.

The p-value for Test 3 is greater than .1. The modeled variance appears to be appropriate here.

The p-value for Test 4 is greater than .1. The model chosen seems to adequately describe the data.

Benchmark Dose Computation

Specified effect = 0.05
 Risk Type = Relative risk
 Confidence level = 0.95
 BMD = 17.6532
 BMDL = 13.4352

Table A8. Parameter Estimates

Parameter Estimates				
Variable	Estimate	Std. Err.	95.0% Wald Confidence Interval	
			Lower Conf. Limit	Upper Conf. Limit
alpha	50.0614	10.4385	29.6023	70.5206
intercept	49.6614	1.66787	46.3924	52.9303
v	-38.2158	3.63855	-45.3472	-31.0843
n	5.69111	1.66023	2.43711	8.9451
k	28.2041	2.95884	22.4048	34.0033

In the *Parameter Estimates* section in the numerical output (Table A8), the parameter estimates associated with the fitted dose-response model are given. Standard errors (Std. Err.) and 95 % confidence intervals for the estimated parameter are also given. Table A8 should not be mixed up with the *Default Initial Parameter Values* given earlier in the numerical output (Table A7), which are the start values for the estimation process. Different types of continuous models (describing the mean response) are available in BMDS, which differ in the number of parameters. The Hill model, which is used in this example, has four parameters. In the case of continuous data not only a model for the mean response, but also a model for the variance needs to be selected. A constant or a non-constant variance model (a power function of the mean response) can be selected in the BMDS. In this example, a constant variance model was selected. The constant variance parameter is called “alpha” in Table A8, while the other parameters given (intercept, v, n, and k) are associated with the Hill model that describes the mean response. So in total, five parameters are here estimated (Table A8).

Table A9. Table of Data and Estimated Values of Interest

Table of Data and Estimated Values of Interest						
Dose	N	Obs Mean	Est Mean	Obs Std Dev	Est Std Dev	Scaled Res.
0	10	51	49.7	8	7.08	0.598
10	9	48	49.6	6	7.08	-0.66
20	10	45	44.9	8	7.08	0.0325
40	9	16	16	9	7.08	-0.02
60	8	12	12	5	7.08	0.0163

In *Table of Data and Estimated Values of Interest* (Table A9) the data set under evaluation is presented, and also the estimates for the mean response and the standard deviation at each dose level (Est Mean, and Est Std Dev).

Table A10. Likelihoods of Interest

Likelihoods of Interest			
Model	Log(likelihood)	# Param's	AIC
A1	-112.603521	6	237.207041
A2	-110.717413	10	241.434825
A3	-112.603521	6	237.207041
fitted	-113.004768	5	236.009537
R	-155.818533	2	315.637067

In *Likelihoods of Interest* (Table A10), a number of log-likelihoods are given. As previously mentioned the log-likelihood can be seen as quantitative measure of how close the model is to the data points. For continuous data, a more extensive comparison of log-likelihoods can be made since both a model for the mean and the variance is established. Several theoretical reference “models” (A1, A2, A3, and R) can thus be defined, which the fitted model for the mean (e.g. the Hill function) and/or the variance (constant or non-constant) can be compared to. The log-likelihoods for five different models (A1, A2, A3, fitted, and R) are calculated for continuous data (in the evaluation of a single data set). These different models, or cases, are described below:

Model A1: Full model with a constant variance; it has one parameter for each dose group that describes the mean, and a single parameter that describes the variance.

Model A2: Full model for both the mean and the variance; it has two parameters for each dose group, one for the mean and one for the variance. Model A2 fits the data perfectly.

Model A3: Full model with a user specified model for the variance. A3 has one parameter for each dose group that describes the mean and one or two additional parameters for the variance. In the case constant variance is selected A3 becomes equivalent to A1.

Fitted model: User specified model/s for the mean response and the variance.

Model R: Null model that represents a “no dose-response relationship”; a horizontal line that corresponds to the average effect over all doses.

Table A11. Test of Interest

Tests of Interest			
Test	-2*log(Likelihood Ratio)	Test df	p-value
Test 1	90.2022	8	<.0001
Test 2	3.77222	4	0.4377
Test 3	3.77222	4	0.4377
Test 4	0.802495	1	0.3703

In *Tests of Interest* (Table A11), the log-likelihood values for the different models (A1, A2, A3, fitted, and R) are compared using likelihood ratio tests. Four different tests, Test 1-4, are considered:

Test 1, R vs. A2: tests if the response and/or variance differ between the dose levels.

Test 2, A1 vs. A2: tests if the variance is homogenous (constant)

Test 3, A3 vs. A2: tests if the selected constant variance model is appropriate.

Test 4, Fitted vs. A3: tests if the selected Hill model is appropriate for the mean response.

To understand better what is done here one should consider what is identical and what is different between the model pairs that are compared in each Test; the aspect that is different is what is “tested”. For example, Test 4 is between model A3 and the fitted model. These models are identical in the variance, but differ in the mean. It is, therefore, testing if the models for the mean differ between “A3” and “fitted” in statistical terms, and if it does not, the fitted model is accepted. The difference between two log-likelihoods for each test pair can be compared with the critical values in Table 4 of the opinion, using the difference in the number of parameters, as the number of Degrees of Freedom (DF).

In Table A11 it is shown that Test 2 and Test 3 give the same p -value. This is because the constant variance model was selected. In this case, model A3 will be the same as model A1. If the non-constant variance model is chosen, the outcomes of Test 2 and 3 will differ.

For Test 1, the BMDS uses a cut-off $p \leq 0.05$. For Test 2-4, however, the BMDS uses a cut-off $p \leq 0.10$ (see Table A7). If the difference between log-likelihood’s associated with two models is so small that it gives a p -value greater than these cut-off’s, the BMDS considers the two models to be statistically similar. As mentioned in the section on quantal data, a cut-off $p \leq 0.05$ was used for all tests. This opinion recommends using a default cut-off $p < 0.05$ for rejecting a model resulting in a greater likelihood of accepting simpler models (see section 5.4).

Table A12. Benchmark dose computation

Benchmark Dose Computation	
Specified effect =	0.05
Risk Type =	Relative risk
Confidence level =	0.95
BMD =	17.6532
BMDL =	13.4352

Finally, in Table A12 the selected value of the BMR (0.05), BMR type (Rel. Dev.), and confidence level (0.95) are given together with the resulting BMD and its lower bound, BMDL. In quoting these values, only two significant figures should be used, i.e. BMDL = 13.

1.5 Advanced features of the BMDS

The illustrations in the sections 1.3 and 1.4 above considered the case when a specific model is fitted to an individual data set. More advanced options in the BMDS allow the user to fit several models to a data set, or fitting a specific model to several data sets, or a combination of both. Below follows a brief guide on how this is carried out using BMDS.

After BMDS 2.0 software has been opened, under “*File*” in the main window of the BMDS, the user selects “*New Session*”. A new window will then appear with a single row and a few columns. Under “*Session Grid*” in the main window, additional rows may be added to the session window. The number of rows added here should correspond to the number of models the user want to fit, and/or the number of data sets that is to be evaluated. In the example below (Figure A4, Table A13), seven models were fitted to the same quantal data set that was used in section 1.1. Thus, seven rows were added to the session window. For each row added to the session window, specifications need to be made with regard to each column “*Model Type*”, “*Model Name*”, “*Data File*”, “*Run*”, “*Model Option File*”, “*Endpoint*”, and “*Out File*”. The options under each column category are described below.

Model Type: This is a drop down menu where the general type of model is selected (see section 1.1 for the available options). In this example, “dichotomous” was selected for each row.

Model Name: A specific model is here chosen from the pool of models that is available for the particular “Model Type” category selected (see section 1.1 for the available options). This is done by right clicking in this column (right click in the cell/s below “Model Name”). In this example, the Weibull, gamma, multistage, log-logistic, logistic, log-probit, and probit models were selected on each row, respectively.

Data File: By right clicking in this column a number of choices regarding the data can be made. “New Data File” can e.g. be selected; a window will then appear into which the user should enter the dose-response data that is to be analysed. This window is identical to that discussed for the simpler analyses in sections 1.3 and 1.4. The data (doses, responses etc) should be entered column-wise:

-In the case of quantal data, the doses, the number of subjects in each group, and the incidence (or fraction responding) in each group should be entered.

-In the case of continuous data, the doses, the number of subjects in each group, the mean response in each group, and the standard deviation in each group should be entered. Alternatively, individual data (individual doses and responses) can be entered.

The columns into which the data was entered may be renamed. This is done by right clicking on the column header. The option “Rename Column” can then be selected. For example, give the name “Dose” for the column into which the doses were entered. After data has been entered and columns named, the data window/file can be saved (in “.dax” format).

Run: Activation of this box simply means that the data set on that particular row will be evaluated.

Model Option File: By right clicking in this column a number choices regarding model options can be made. “New Option File” can e.g. be selected; a window will then appear in which the user needs to make some specification:

Column Assignments:

It needs to be specified what each column in the previously added data file represents.

Optimizer Assignments:

The default iteration settings are here given. These values normally do not need to be changed by the user.

Parameter Assignments:

The default start values for the model fitting process, with regard to each parameter in the selected model, are here given. These values normally do not need to be changed by the user.

Other Assignments:

A number of options can be activated here depending on the model selected. Activation of the box “BMD calculation” simply means that the BMD and BMDL will be calculated. The model specific restrictions described in section 1.1 can also be activated here. The confidence level for the BMDL calculation should be specified and also the BMR, and the BMR Type (see section 1.2 for the available options). The user may also specify the adverse direction of the dose-response; “up”, or “down”, or leave this to the software; “automatic”. In this example, confidence level = 0.95, BMR = 0.10, and BMR Type = extra, was selected on each row (for each model). After the different choices have been made this file can be saved (in “.opt” format).

Endpoint: Nothing needs to be done here. The name given by the user for the column in the data window/file (that was added under “Data File”) where the incidences (quantal data) or mean responses (continuous data) were entered will be displayed here.

Out File: By right clicking on this column “Set Out File To” can e.g. be selected which allows the specification of a file name for the output results (in “.out” format). The output results can also be copied and saved after it has been generated.

After including all relevant information in the steps above, the session window can be saved (in “.ssn” format). The session is executed by pressing “Run” in the session window.

The output from the example discussed (with the seven models fitted to the quantal data set that was also used in section 1.3) is illustrated in Figure A4 and Table A13. Figure A4 shows the seven models fitted to the data, and Table A13 gives a summary report of the numerical output. The information in Table A13 is similar to that discussed in section 1.3, but is given here in summary form with regard to each model. The complete numerical output can also be derived. This is done by right clicking in any cell of the particular column, i.e. model, of interest in the summary report, and select “Show Out/Graph”. A new window will then appear consisting of two parts: the plot for the particular model considered and a complete numerical output. The output for the model of interest is given in a format identical to that discussed in sections 1.3 and 1.4.

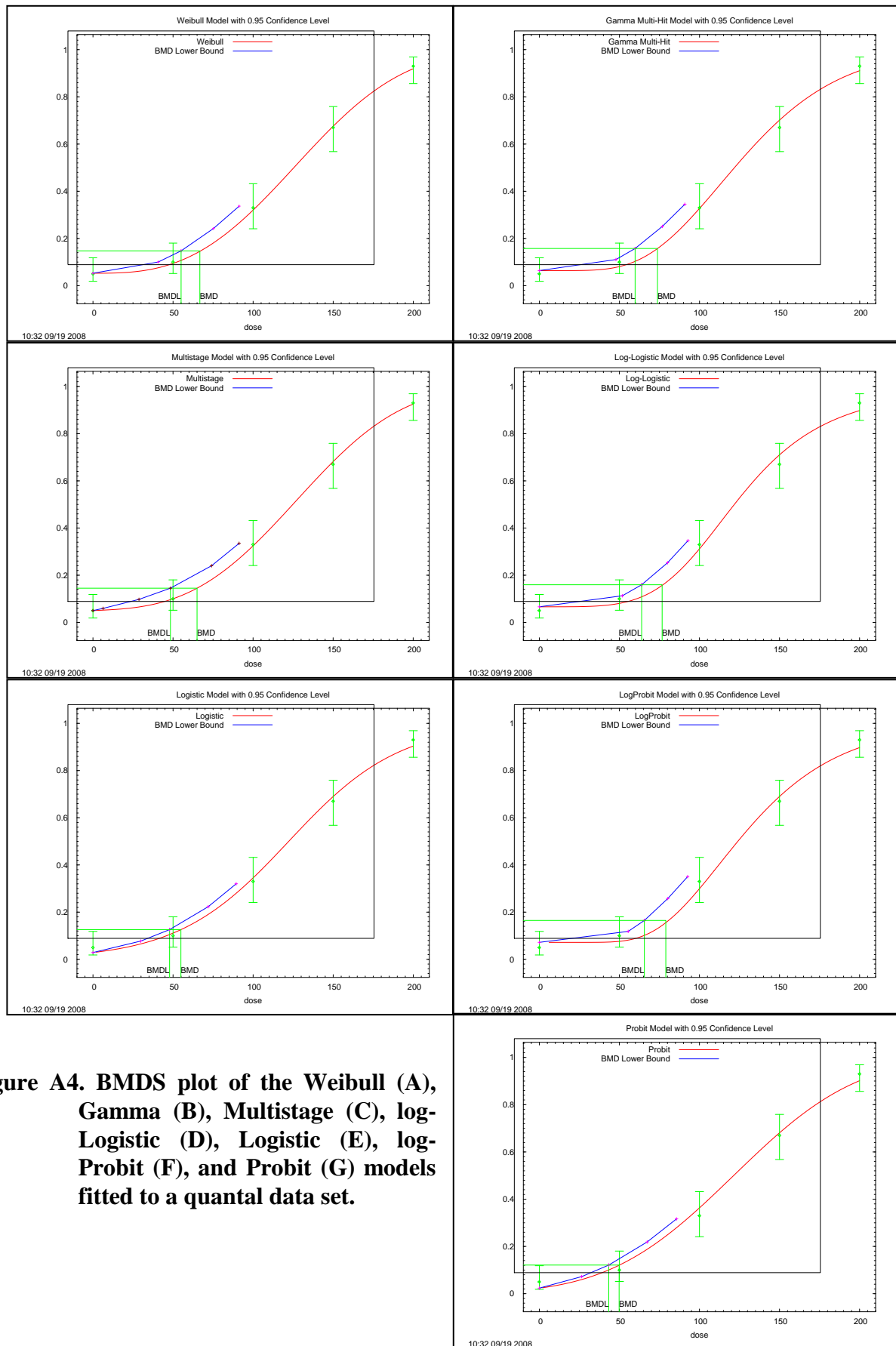


Figure A4. BMDs plot of the Weibull (A), Gamma (B), Multistage (C), log-Logistic (D), Logistic (E), log-Probit (F), and Probit (G) models fitted to a quantal data set.

Table A13. BMDS Summary report of numerical output from seven quantal dose-response models.

Model Name	Weibull	Gamma	Multistage	LogLogistic	LogProbit	Logistic	Probit
Data File Name	1-23-Dicho.dax	1-23-Dicho.dax	1-23-Dicho.dax	1-23-Dicho.dax	1-23-Dicho.dax	1-23-Dicho.dax	1-23-Dicho.dax
Option File Name	T.opt	T.opt	T.opt	T.opt	T.opt	T.opt	T.opt
The parameter betas			are restricted to be positive				
Dependent variable	EFFECT2	EFFECT2	EFFECT2	EFFECT2	EFFECT2	EFFECT2	EFFECT2
Independent variable	DOSE	DOSE	DOSE	DOSE	DOSE	DOSE	DOSE
Maximum number of iterations	250	250	250	250	250	250	250
Relative Function Convergence has been set to	1e-008	1e-008	1e-008	1e-008	1e-008	1e-008	1e-008
Parameter Convergence has been set to	1e-008	1e-008	1e-008	1e-008	1e-008	1e-008	1e-008
Total number of Dose Groups with missing values	0	0	0	0	0	0	0
Slope parameter				is restricted as slope ≥ 1	is restricted as slope ≥ 1	is not restricted	is not restricted
Total number of observations	5	5	5	5	5	5	5
Total number of parameters in model			4				
Total number of specified parameters			0				
Degree of polynomial			3				
Initial/Specified Background	0.0544554	0.0544554	0.0542477				
Initial/Specified Slope	4.35787e-007	0.0695017		3.75823	2.13839	0.0272189	0.0158216
Initial/Specified Power	parameter is restricted as power ≥ 1	parameter is restricted as power ≥ 1					
Initial/Specified Intercept				-17.8385	-10.1439	-3.21894	-1.87282
Initial/Specified Beta(1)			0				
Initial/Specified Beta(2)			0				
Initial/Specified Beta(3)			3.23694e-007				
Asymptotic Correlation Matrix of Parameter Estimates	Array	Array	Array	Array	Array	Array	Array
Parameter Estimates	Array	Array	Array	Array	Array	Array	Array
Analysis of Deviance Table	Array	Array	Array	Array	Array	Array	Array
AIC	415.312	416.885	415.227	418.198	418.094	415.73	417.468
Goodness of Fit	Array	Array	Array	Array	Array	Array	Array
Chi-square	0.19	1.75	0.11	3.00	2.93	2.82	4.97
DF	2	2	2	2	2	3	3
P-value	0.9078	0.4167	0.9477	0.2232	0.2314	0.4197	0.1737
Specified effect	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Risk Type	Extra risk	Extra risk	Extra risk	Extra risk	Extra risk	Extra risk	Extra risk
Confidence Level	0.95	0.95	0.95	0.95	0.95	0.95	0.95
BMD	66.7396	73.7378	65.0092	76.6894	78.9806	54.9049	49.8127
BMDL	55.0227	59.7761	48.3869	63.8925	65.5908	47.8345	43.3441
BMDU			72.3484				

Note: For the Multistage model the BMDS also computes the upper bound BMDU; a two-sided 90% confidence interval is here given for the BMD.

2 The PROAST software

PROAST can be downloaded from the RIVM website (www.rivm.nl/proast). Full details for installation and a step-by-step guide to its use are provided in the documentation that comes with the software.

PROAST is written in the S language, and it is, therefore, necessary to have one of the software package Splus or R available: PROAST is a module that is added to either the Splus or the R software. A license needs to be purchased for Splus, while the R software can be downloaded free of charge from the R-website www.cran.r-project.org. An advantage of Splus is that it provides better graphical output. However, the following illustration will use the R software as a basis, because this software is more easily accessible.

The R version of PROAST is designed in terms of a set of multiple choice questions that guide the user through the different options available. For example, different options are given for what type of data the user wants to analyze, e.g. continuous or quantal data? Does the user want to fit a specific model function, or a family of nested models? And what BMR should be selected? In a similar manner, PROAST also needs to gather details regarding the data; e.g. what parts in the data matrix represent the “dose” and the selected “response”? The option to include covariates exists, e.g. does the user want the analysis to be stratified over two sub-groups (e.g. males and females)? Different options are given on how to plot the data and the fitted model including an option for the scales used.

An analysis of a continuous dataset by PROAST is illustrated below. In this analysis the exponential family of models was specified with a BMR of 5%. A numerical output of the analysis is generated (Table A14) together with three plots (Figures A5, A6, and A7).

In Table A14, the resulting log-likelihood for each model in the exponential family (models m11, m1, m2, m3, m4, and m5) is given. As previously pointed out, the log-likelihood is a quantitative measure of how well the respective model fits the data. The number of parameters (npar) for each model is indicated, and information is given about the model fitting process (converged); if the value in this column reads “1” it means that the model fitting process was successful, i.e. convergence was achieved; not necessarily that a good model was achieved.

Table A14. Numerical output generated in PROAST.

```

----- relKidney -----
ATTENTION: There are 1 missing values

No observations below detection limit

ANALYSIS WITH EXPONENTIAL MODELS

model          converged      npar   loglik
m11-           1                8      59.07
m1-            1                2      32.3
m2-           1                3      49.37
m3-           1                4      49.37
m4-           1                4      56.72
m5-           1                5      58.02
re-fitting selected model m4-
m4-           1                4      56.72

calculating C.I.....

the CED (in orig. units) and the 90 % confidence interval for group 1 is:
47.102
30.57046
76.4217

The number of parameters in this output includes the variance parameter.

```

Model m11, the full model fits the data points perfectly. The number of parameters for model m11 is 8, which equals the number of dose groups (7), including control, plus one parameter for the variance.

Model m1, the null model represents a “no dose-response relationship”; a horizontal line that corresponds to the average effect over all doses.

Models m2, m3, m4, and m5 represent the user-specified models, i.e. the nested models of the exponential family (see Table 3 of the opinion).

Figure A5 shows a graphical comparison of all the fitted models.

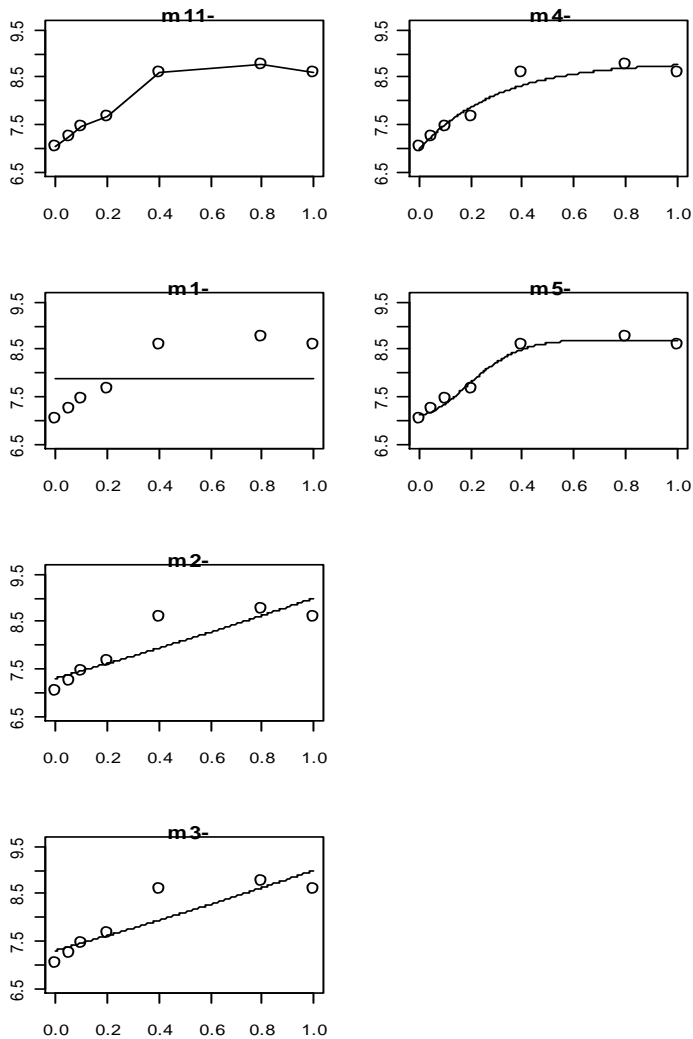


Figure A5. Graphical comparison of the exponential models fitted by PROAST.

The circles represent the observed means at different doses. The x-axis reflects doses; the y-axis the response. For the equations of the models, see Table 3

When a family of models is selected by the user to be fitted, PROAST operates so that it determines the most appropriate model for the particular data considered. This is done by pairwise comparison of the models using likelihood ratio tests, on the significance level $p \leq 0.05$.

In the example, m4 is determined to be the most appropriate model. The selected model is shown labelled “re-fitting selected model” together with the BMD (47) which is called Critical Effect Dose (CED) in the numerical output and its two-sided 90% confidence interval. The lower bound of the interval, the BMDL, is 31.

Figure A6 shows a detailed plot of the data and the selected model (m4).

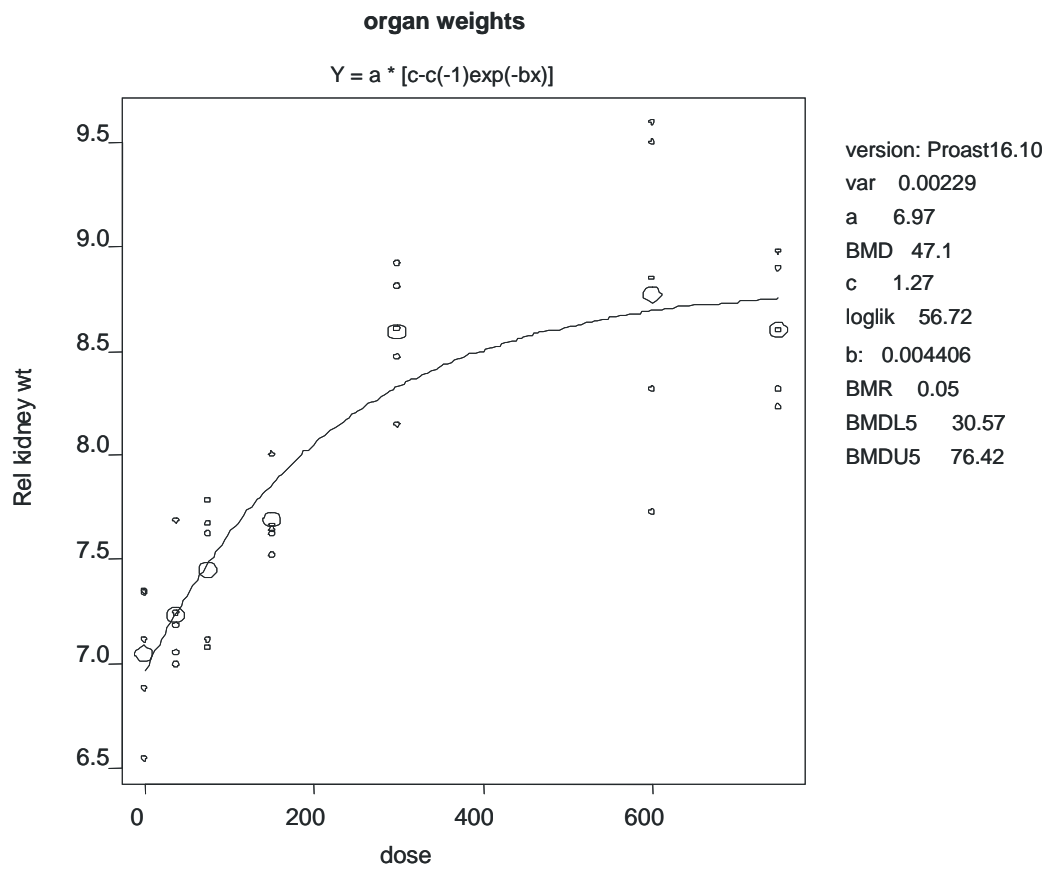


Figure A6. Plot of the data and the selected model (m4) in PROAST. The specified BMR, the BMD, and the BMDL are shown.

Figure A7 shows the log-likelihood profile computed by PROAST. The value of the BMD (47, which is 1.67 on log-scale) corresponds to the maximum of the log-likelihood. The upper and lower 95% confidence bounds for the BMD are the doses where the profile curve intersects with the horizontal line in Figure A7; this line corresponds to a log-likelihood (on the y-axis) which is on the border of being significantly different from the maximum log-likelihood at the significance level $p \leq 0.10$, given the particular model (m4 with constant variance) and data considered. This is equivalent to a two-sided 90% confidence interval.

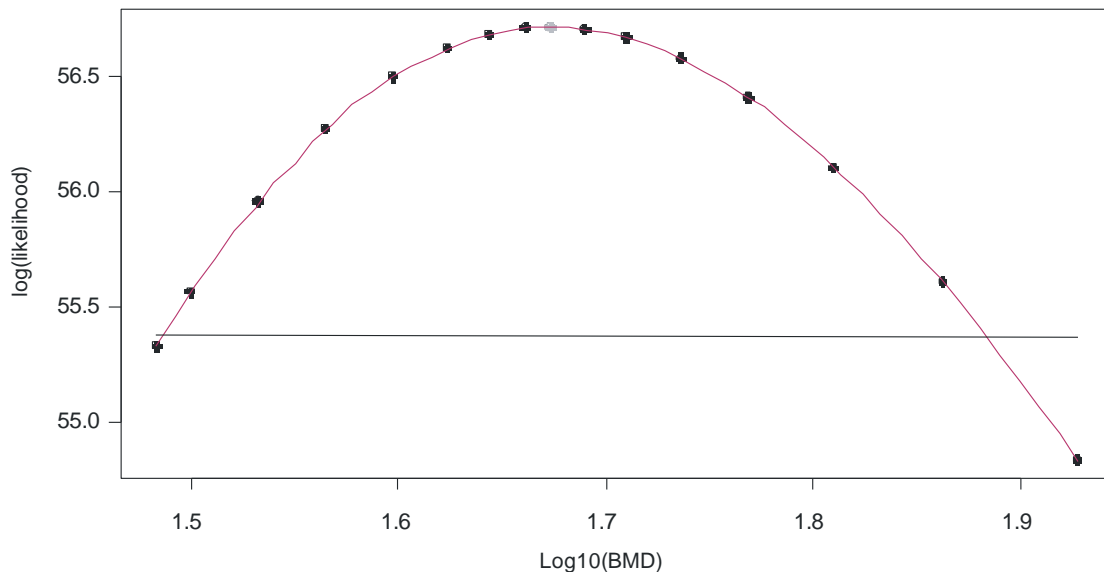


Figure A7. Plot generated by PROAST of the log-likelihood profile.

PROAST also offers a number of additional options, e.g. use of covariates, which are not described here. Those interested are referred to the PROAST documentation.

3 Summary of differences between the BMDS and PROAST

A comparison of the BMDS and PROAST software is provided in Table A15.

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

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

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

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

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

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

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

Table A15. Comparison of BMDS and PROAST

	BMDS	PROAST
Environment	Can be run immediately (as an executable) under Windows	Splus or R is required. R is free software
First use	Easy to get started	Higher threshold; requires basic understanding of Splus or R
User interaction	Fully Windows-based	Ordered process of answering multiple choice questions; Graphical User Interface available only for continuous data
Continuous data	Yes	Yes
Nested continuous data, e.g. for litter effects	No	Yes
Quantal data	Yes	Yes
Nested quantal data, e.g. for litter effects	Yes	Yes
Ordinal data	No	Yes
Confidence interval by profile likelihood	Yes	Yes
Confidence interval by bootstrapping	No	Yes
Covariates	No	Yes
Automatic model fitting for nested models	Yes (beta version)	Yes
Graphical output	Yes	Yes