Review
The current state of knowledge on the use of the benchmark dose concept in risk assessment

Salomon Sand,* Katarina Victorin and Agneta Falk Filipsson
Institute of Environmental Medicine, Karolinska Institutet, P.O. Box 210, SE-17177 Stockholm, Sweden

Received 20 February 2007; Revised 13 June 2007; Accepted 17 July 2007

ABSTRACT: This review deals with the current state of knowledge on the use of the benchmark dose (BMD) concept in health risk assessment of chemicals. The BMD method is an alternative to the traditional no-observed-adverse-effect level (NOAEL) and has been presented as a methodological improvement in the field of risk assessment. The BMD method has mostly been employed in the USA but is presently given higher attention also in Europe. The review presents a number of arguments in favor of the BMD, relative to the NOAEL. In addition, it gives a detailed overview of the several procedures that have been suggested and applied for BMD analysis, for quantal as well as continuous data. For quantal data the BMD is generally defined as corresponding to an additional or extra risk of 5% or 10%. For continuous endpoints it is suggested that the BMD is defined as corresponding to a percentage change in response relative to background or relative to the dynamic range of response. Under such definitions, a 5% or 10% change can be considered as default. Besides how to define the BMD and its lower bound, the BMDL, the question of how to select the dose-response model to be used in the BMD and BMDL determination is highlighted. Issues of study design and comparison of dose-response curves and BMDs are also covered. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: benchmark dose; BMD; BMDL; risk assessment; safety assessment

Introduction
Risk assessment involves characterizing and quantifying the potential adverse effects on humans of exposure to radiation, other physical agents, chemicals, or situations that may constitute a human health or environmental hazard (McClellan, 1999). Risk assessment of chemicals has two major objectives; either to assess the health risks associated with a particular type of exposure or to establish acceptable exposure levels for humans.

Risk assessment of chemicals has traditionally been performed differently for genotoxic carcinogenicity compared with non-genotoxic endpoints. For genotoxic carcinogens it is considered that even very low levels of exposure may increase the risk for adverse outcomes. This assumption is based on the idea that there is a theoretical possibility that exposure to a single molecule of a genotoxic substance may be sufficient to initiate the development of a tumor (USEPA, 1995; Edler et al., 2002). Because of this, for genotoxic carcinogens, risk is traditionally assumed to be proportional to dose at low doses. Several dose-response models have been suggested for low dose extrapolation in risk assessment of genotoxic carcinogens (Krewski and Van Ryzin, 1981; Edler and Kopp-Schneider, 1998). However, the policy-based risk levels considered acceptable for the human population are orders of magnitudes lower than those observed experimentally; risk levels in the range of $10^{-5}$ to $10^{-6}$ are usually discussed (Edler et al., 2002). Since different models can provide very different results in the low dose region (Krewski and Van Ryzin, 1981) it has been considered controversial to explicitly rely on any of the models available. Procedures of linear extrapolation from the region of observable response have instead been suggested. The current guidelines for risk assessment of genotoxic carcinogens, developed by the United States Environmental Protection Agency (USEPA), do not show any preference for a certain model, rather it is suggested that a number of dose-response models can be considered for estimation of the lower bound of a dose corresponding to risk levels of 1%, 5% or 10%. Any of these values may then be considered as a point of departure, typically subject to linear extrapolation (USEPA, 2005). Recently, the margin of exposure approach, rather than linear extrapolation, has been suggested for risk assessment of substances that are both genotoxic and carcinogenic (Barlow et al., 2006). The use of dose-response modeling has also been recommended for derivation of the starting point (the lower bound of a dose corresponding to, for example, a 10% increase in tumor incidence over background) in the margin of exposure approach (Barlow et al., 2006).
For the majority of toxicological effects, excluding genotoxic effects, it is assumed that there exists an exposure threshold below which no adverse effect will occur. It has been argued that in order to produce a toxicologically relevant response a high enough concentration of the chemical is required to exceed the homeostatic and cytoprotective processes (Dybing et al., 2002; Edler et al., 2002). The traditional approach in health risk assessment of non-genotoxic agents is based on the threshold assumption and involves establishment of a no-observed-adverse effect-level (NOAEL). The NOAEL is usually derived from animal data and is defined as the highest experimental dose level for which the response is not significantly different compared with the response in the control group (see Figure 1). The NOAEL is traditionally used as a starting point in the determination of health based guidance values such as acceptable daily intakes (ADIs) by the application of uncertainty factors (WHO, 1999).

Serious criticism has been raised towards the NOAEL approach over the years and the benchmark dose (BMD) method has been suggested as an alternative. The BMD concept was introduced for human health risk purposes in the mid-1980s (Crump, 1984), but methodological and practical issues of this approach are still discussed in the scientific literature. To date, the BMD method has mostly been employed by authorities, such as the EPA, and consultants in the USA while it has not been considered to the same extent in Europe. Research specifically related to the BMD approach has been performed at only a few institutions in Europe, and practical as well as theoretical knowledge of the method is limited on an overall basis. Awareness of the BMD concept has, however, increased, for example through its application in some EC research projects, and the method is also discussed within the European Food Safety Authority (EFSA). In general, there is an increasing focus on statistical methods in the area of quantitative risk assessment within the European Community.

The purpose of this review is to report on the scientific state of knowledge regarding the BMD concept. Emphasis is placed on discussing the different procedures that have been suggested for BMD analysis and describing how the BMD method has been applied for different types of response data (quantal and continuous data). Important questions that need consideration are highlighted. These include how to define the BMD and how to select a dose-response model for BMD determination. Aspects of study design in dose-response modeling and comparison of dose-response curves and BMDs are also discussed.

**Overview of the Benchmark Dose Approach**

**Background**

The BMD concept was introduced by Crump (1984) and involves fitting a mathematical model to dose-response data. The BMD is defined as the dose causing a predetermined change in response. This specified response change is generally referred to as the benchmark response (BMR). The lower confidence limit of the BMD, i.e. the BMDL, has been proposed to replace the NOAEL as a starting point for determination of health based guidance values (i.e. uncertainty factors are applied to the BMDL instead of to the NOAEL). The BMDL is usually defined as the one-sided lower 95% confidence limit on the BMD (which equals the lower bound of a two-sided 90% confidence interval); it can be interpreted as the dose corresponding to a response not likely to be larger than the specified BMR (with 95% confidence). The BMD method is illustrated in Figure 1.

Even though some early attempts were made to generalize the BMD concept (Crump, 1984; Gaylord and Slikker, 1990), initial discussions mostly concerned its application to quantal data (incidence data) with the focus towards developmental toxicity (Faustman et al., 1994; Allen et al., 1994a, 1994b; Kavlock et al., 1995). The BMD with lower bound, BMDL, was originally presented as the dose causing an excess risk of 1–10% (Crump, 1984; Kimmel and Gaylord, 1988). Using a number of illustrative examples,
Crump (1984) argued that according to this definition the BMDL will not depend strongly upon the model selected since the procedure does not involve extrapolation far below the experimental range. Apparently, the current U.S. EPA guidelines for determination of the point of departure in risk assessment of genotoxic carcinogens are identical to the original definition of the BMD (and BMDL) (USEPA, 2005).

A number of arguments for use of the BMDL rather than the NOAEL have been presented in the scientific literature (Crump, 1984; Kimmel and Gaylor, 1988; Leisenring and Ryan, 1992; Allen et al., 1994a; Barnes et al., 1995; USEPA, 1995). The fact that the BMD method involves uncertainty analysis, by the establishment of a confidence interval and use of the BMDL, has been put forward as a major improvement relative to the NOAEL approach. As the sample size becomes higher, confidence intervals tend to be shorter, reflecting a lower uncertainty, which possibly will result in a higher BMDL; the BMDL increases with increasing sample size considering all else equal. Conversely, the fact that the NOAEL tends to decrease with increasing sample size has been discussed as an inappropriate property in a regulatory setting; larger experiments and more information should possibly allow the guidance value to be higher, and not the opposite (Crump, 1984; Leisenring and Ryan, 1992). Other important and commonly discussed advantages of the BMD concept include that it makes more use of all the available data which results in the shape of the dose-response being considered to a higher extent; a starting point for risk assessment derived using the BMD method is not limited to being one of the experimentally selected dose levels; a BMDL can often be calculated even on occasions when a NOAEL is not present in the study; and while the response at the NOAEL (in terms of BMR equivalent) is not assessed and may vary from case to case, the BMDL corresponds to an explicit response level which introduces consistency.

In spite of these advantages there has been a reluctance to use the BMD approach (Travis et al., 2005). The BMD method is a more advanced approach and competence in the area is limited. The further need for harmonization with respect to some aspects of the BMD concept (e.g. how the BMD should be defined) may also have prevented it from being generally accepted. The question of how to apply the BMD concept in the case of very limited data has also been discussed. Such situations are not ideal for risk assessment but sometimes a reality. In the NOAEL approach routines are established for how to deal with situations related to this issue; e.g. in the case where no significant effects are observed the highest dose can be regarded as the NOAEL, and in the case where only significant effects are observed the lowest-observed-adverse-effect level (LOAEL) is divided by a factor to give a ‘surrogate’ NOAEL. The fact that such data include limited information about the dose-response becomes more evident when the BMD method is applied. For example, in the case of no significant effects there may not be any support for a dose-response trend and a BMDL may not be derived, suggesting that data are too poor. Alternatively, in the case where a NOAEL is not present and where the response at the LOAEL is clearly higher than the benchmark response, the BMD method involves extrapolation, which potentially can result in large discrepancies between BMDLs from different models that adds to the uncertainty. However, in less extreme cases the BMD method will not result in substantial extrapolation even though a NOAEL is not present.

Basic Issues in BMD Analysis

Many of the issues of how to conduct a BMD analysis concerns those associated with dose-response modeling in general; how to describe the data with a mathematical model and how to derive a BMD (or an effective dose, ED) from that model. A number of decisions need to be made in this respect.

Fitting Method

Fitting a dose-response model to data involves finding optimal values of the parameters (or variables) that specifies the model. Using a search algorithm the values of the parameter are varied until a general fit criterion is fulfilled; the model is then aligned as close as possible to the data points, given the fit criterion used. In the BMD field the default fit criterion is to maximize the so-called likelihood; this approach is referred to as the maximum likelihood method. Under this method a distribution for the data needs to be assumed; a quantal response is assumed to be binomially distributed and a continuous response is generally assumed to be normally or log-normally distributed.

Definition of the BMD and BMDL

An appropriate definition of the BMD needs to be specified. This relates to what change in response the BMD should correspond to and also how a change in response generally should be defined (e.g. % change relative to background). Several BMD definitions have been presented and they are discussed in detail later. The lower bound of the BMD, the BMDL, needs to be defined and estimated using a selected method for confidence interval calculation. The BMDL is generally defined as the one-sided lower 95% confidence limit of the BMD (or alternatively, the lower bound of a two-sided 90% confidence interval). The likelihood ratio test† is typically the method applied for estimating the BMDL (Moerbeek et al., 2004).
Selection of Dose-response Model and BMDL

Several dose-response models may be fitted to a given data set. In the field of BMD analysis, models are used descriptively and it is assessed in a statistical sense how close the fitted model is to the data points; a model is typically rejected if \( p \leq 0.05 \). In cases where several models are not rejected, further considerations of model choice for BMDL determination are needed. Acceptable models may be further compared based on statistical analysis; for example using likelihood ratio tests\(^\dagger\), or the model providing the most conservative BMDL can be selected. Individual dose-response models and model selection are discussed in more detail later.

Detailed Review of the Benchmark Dose Method

Data Format

Data used for dose-response modeling and BMD calculations may either be in a continuous or quantal format. While the BMD concept was quite straightforwardly introduced for quantal data its application to continuous dose-response information has merited more discussion (Barnes et al., 1995; Kavlock et al., 1995; Gaylor et al., 1998; Crump, 2002; Slob, 2002; Falk Filipsson et al., 2003; Gaylor and Aylward, 2004; Sand et al., 2003, 2004, 2006). In contrast to quantal data where experimental animals are categorized as responders or non-responders, for continuous data the degree of response is observed in the individual subject. Typical continuous responses constitute changes in organ weights and enzyme activities. Dose-response data may also be presented in ordinal format. In this case, several sub-classes describing the severity of response are defined; responses in subjects may, for example, be classified as mild, moderate or severe. In its most simple case ordinal data are equivalent to quantal data. And in its most extreme, or theoretical, case it becomes equivalent to continuous data, i.e. when the number of response classes are infinite. Because of this, quantal data (e.g. the fraction of animals classified as having atrophy) may be connected to an ‘underlying’ continuous response (e.g. the degree of atrophy) (Slob and Pieters, 1998).

\(^\dagger\) Likelihood ratio tests may be used to compare model fits and construct confidence intervals. It can be shown that the test statistics, \( -2(\ln L_1 - \ln L_2) \), where \( \ln L_1 \) and \( \ln L_2 \) is the log-likelihood under model 1 and 2 respectively, approximately follows a \( \chi^2 \) distribution with degrees of freedom equal to the difference in the number of parameters between the two models. A critical level \( \alpha = 0.05 \) is commonly employed as default in this type of testing, i.e. if \( p \leq 0.05 \) the two models are considered to be significantly different. For establishment of the BMDL (the one-sided lower 95% confidence limit of the BMD) a critical level \( \alpha = 0.10 \) is used, rather than \( \alpha = 0.05 \).

In the scope of originally defining the BMD as the dose corresponding to some level of risk for adverse effects (for incidence data), implementation of risk (or probability) based procedures have also been discussed for continuous dose-response information (Gaylor and Slikker, 1990; Kodell and West, 1993; Crump, 1995). Alternatives to such formulations have also been developed where the BMD is estimated from the mean response function (Crump, 1984; Murrell et al., 1998; Slob and Pieters, 1998; Sand et al., 2006). An example of the latter approach to apply the BMD concept to continuous data is illustrated in Figure 2. The issues with continuous data will be discussed in detail later.

Figure 2. An illustration of how the BMD approach can be applied to continuous endpoints. Data observed in Han/Wistar rats following long-term exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) is used as an example (data from Viluksela et al., 2000). Continuous responses may either increase or decrease with increasing dose; an example of a decreasing response is body weight gain. A dose-response model (Hill function, Table 2) has been fitted to the mean responses (triangles). The BMD is defined as corresponding to a 10% change in response relative to the dynamic range of response (the difference between the maximum and minimum response level, predicted by the fitted curve). In this particular example, the BMD also corresponds to a 5% change in response relative to the background response (predicted by the model) at dose zero. This relation occurs due to the fact that the fitted model levels off at a response around 45, which is half the estimate of the background response of 90. The BMDL is illustrated by the dashed–dotted line and represents the one-sided lower 95% confidence limit on the BMD. The dashed curve illustrates part of the lower confidence bound of the fitted curve, which intersects with the BMDL at the specified response (a similar curve may also be established in Figure 1). Several BMD definitions have been presented for continuous data and they are discussed in detail in the text. This figure is available in colour online at www.interscience.wiley.com/journal/jat
Quantal Dose-Response Information

Dose-Response Models

For quantal data, a number of dose-response models have been suggested in health risk assessment of chemicals. These models are given in Table 1. Some of the models are standard probability distribution functions, i.e., the logistic, the log-logistic, the probit, the log-probit and the Weibull models, which are based on the notion that each animal in the population has its own tolerance to the chemical (Krewski and Van Ryzin, 1981). The multi-stage model and the gamma model are considered to be stochastic models that are based on the notion that a positive response in an animal is the result of the random occurrence of one or more biological events (Krewski and Van Ryzin, 1981; Casarett and Doull, 1996). In spite of this difference in traditional interpretation all these models are presently used on a descriptive basis; for example the multi-stage model is not per se regarded as more appropriate compared with the other models in risk assessment of genotoxic carcinogens (USEPA, 2005).

While the models in Table 1 may be used for general quantal response data, they are not designed to handle the special features of data from developmental toxicity testing. In developmental toxicity testing, information regarding the presence or absence of response (for example, a certain type of malformation) in offspring after maternal exposure to a toxic agent is recorded. For such data intralitter correlations may occur, meaning that responses in offspring are more similar within litters compared with between litters (Rai and Van Ryzin, 1985). This correlation problem has prompted the development of methods and models specifically designed for developmental toxicity data (Rai and Van Ryzin, 1985; Kupper et al., 1986; Kodell et al., 1991). In addition to methods for modeling of single endpoints, extensions of how different developmental endpoints may be analysed simultaneously have been presented (Krewski and Zhu, 1994, 1995).

Table 1. Dose-response models for quantal data

<table>
<thead>
<tr>
<th>Model</th>
<th>Weibull model:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(d) = \frac{1}{1 + e^{-\alpha d}}$</td>
<td>$p(d) = \gamma + (1 - \gamma) (1 - e^{-\beta d})^\gamma$</td>
</tr>
<tr>
<td>log-logistic model:</td>
<td>$p(d) = \gamma + (1 - \gamma) \frac{1}{\alpha} \int_0^{\beta d} x^{-\alpha} e^{-x} dx$</td>
</tr>
<tr>
<td>probit model:</td>
<td>$p(d) = \gamma + (1 - \gamma) \frac{1}{\alpha} \int_0^{\beta d} x^{-\alpha} e^{-x} dx$</td>
</tr>
<tr>
<td>log-probit model:</td>
<td>$p(d) = \gamma + (1 - \gamma) \frac{1}{\alpha} \int_0^{\beta d} x^{-\alpha} e^{-x} dx$</td>
</tr>
</tbody>
</table>

Note: In the models, $d$ represents the dose; $\gamma$ is a parameter that describes the background ($\gamma$ does not exist in the logistic and the probit models); $\alpha$ and $\beta$ are additional parameters; and $n$ is the polynomial degree in the multi-stage model.

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

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

BMD Definitions for Quantal Data and Their Applications

As previously stated, the BMD with its lower bound, BMDL, was initially defined as the dose causing a 1–10% increase in risk over background (Crump, 1984). In this context, two slightly different definitions of the benchmark response level (BMR) associated with the BMD have been presented. The BMR may be expressed in terms of ‘additional’ or ‘extra’ risk:

$$BMR = p(BMD) - p(0), \text{ additional risk,}$$

$$BMR = \frac{p(BMD) - p(0)}{1 - p(0)}, \text{ extra risk,}$$

where $p(BMD)$ is the probability of response at the BMD, and where $p(0)$ is the background probability of response. Definitions (1) and (2) coincide when the background equals zero.

Much of the initial discussion of the BMD approach focused on its applicability for developmental toxicity data. In depth investigations of this issue have been presented in a series of articles (Faustman et al., 1994; Allen et al., 1994a, 1994b; Kavlock et al., 1995). In these studies the BMD method was evaluated and compared with the NOAEL approach using a data base containing 246 developmental toxicity experiments representing 1825 endpoints relating to dead implants or malformed fetuses (Faustman et al., 1994). Analysis of the data was performed using both quantal and continuous response definitions. According to initial investigations, a quantal response variable was established using a litter-based approach; i.e. a litter was considered to be affected (or ‘responding’) if one or more fetuses had the endpoint of interest (Faustman et al., 1994; Allen et al., 1994a). This response was analysed using the Weibull model. BMDLs (the lower 95% confidence bound of the BMD) corresponding to additional risks of 1%, 5% and 10% were found to be lower than the NOAEL on average, and the BMDL associated with the 10% risk level most resembled the NOAEL. Models specifically designed for developmental toxicity testing were later applied. Using these models, a BMDL corresponding to an additional risk of...
5% was close to the NOAEL on average (Allen et al., 1994b).

The observations that a BMDL corresponding to an additional risk of 5–10%, on average, seems to be close to the NOAEL for quantal data and developmental toxicity (Allen et al., 1994a, 1994b) have later been used as a basis for the BMR selection. This was exemplified in Allen et al. (1996, 1998) where BMDLs were calculated for boric acid and isopropanol, respectively, based on developmental data observed in rats. More recently, the 5% level was used in analysis of developmental data observed in rats after exposure to dimethoate (Reiss and Gaylor, 2005), which was also based on the observational findings in Allen et al. (1994b) and suggesting that the limit of detection for the typical developmental test is about a 5% incidence (Reiss and Gaylor, 2005).

It has been pointed out that the recommendation of a 5% risk level for quantal data, which has been based on resemblance to the NOAEL, on average, may not constitute a scientific rationale for BMR selection, but rather reflect the level of risk that has been associated with applications of the NOAEL approach in the past (Allen et al., 1994a; Barton and Das, 1996; Setzer and Kimmel, 2003). Also, while comparison between BMDLs and NOAELs has been conducted for developmental toxicity based on a large data base (Faustman et al., 1994), such analysis is more limited with respect to other toxicity endpoints (Barton and Das, 1996; Haber et al., 1998). Haag-Grönlund et al. (1995) performed investigations of effects on liver, kidney, the central nervous system and tumors using trichloroethylene as a model substance. They concluded that the risk at the NOAEL, for a considerable amount of all NOAELs, could be 10% or higher. Using the same data base similar findings have been presented (Falk Filipsson and Victorin, 2003). Moreover, Fowles et al. (1999) compared BMDLs and NOAELs for 120 acute inhalation lethality data sets. In accordance with previous investigations for developmental toxicity (Allen et al., 1994a), BMDLs corresponding to additional risks of 1–10% were, on average, lower that the NOAEL (and BMDL on was, on average, closest to the NOAEL). However, for the acute inhalation data the differences between BMDLs and NOAELs were smaller.

While resemblance to the NOAEL has been used as the most frequent argument for selecting the BMR for quantal data, the 5–10% has also been supported in investigations of how low the BMR may be specified without the BMDL becoming dependent on the choice of dose-response model (Sand et al., 2002). In Sand et al. (2002) all the typical models in Table 1 were applied to malformation data observed in mice fetuses following maternal exposure to polychlorinated and polybrominated compounds. An extra risk of 5% was suggested since the choice of model, for BMDL calculation, did not appear to be important down to this risk level (Figure 3).

### Table 2. Statistical models for continuous data

<table>
<thead>
<tr>
<th>Dose-response models</th>
<th>Variance models</th>
</tr>
</thead>
<tbody>
<tr>
<td>polynomial model:</td>
<td>constant model:</td>
</tr>
<tr>
<td>( \mu(d) = \alpha + \beta_1d + \beta_2d^2 + \ldots + \beta_n d^n )</td>
<td>( \sigma(d) = c )</td>
</tr>
<tr>
<td>power model:</td>
<td>power function of the mean:</td>
</tr>
<tr>
<td>( \mu(d) = \alpha + \beta d^\theta )</td>
<td>( \sigma(d) = \lambda \mu(d) d^\theta )</td>
</tr>
<tr>
<td>Hill function:</td>
<td>dose dependent exponential model:</td>
</tr>
<tr>
<td>( \mu(d) = \alpha + \theta \frac{d^\beta}{\kappa + d^\beta} )</td>
<td>( \sigma(d) = e^{\lambda \mu(d)} (1 + \eta) )</td>
</tr>
<tr>
<td>exponential model:</td>
<td></td>
</tr>
<tr>
<td>( \mu(d) = \alpha + \theta \left[ 1 - e^{-\frac{d^\beta}{\kappa}} \right] )</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** In the models, \( d \) represents the dose; and \( \alpha \) is a parameter that describes the background response. The degree of the polynomial model is denoted by \( n \). In the Hill function and the exponential model, \( \theta \) is a parameter that describes the dynamic range of response; \( \kappa \) is the location parameter (which equals the ED50 in the Hill model); and \( \eta \) is a parameter that describes the shape of the dose-response curve. The Hill function is symmetrical on the log-dose scale. Note that parameters \( \kappa \) and \( \eta \) do not have identical interpretation across the different models. This is also the case for parameters \( \lambda \) and \( \rho \) that define the non-constant variance models.
Figure 3. Ratios between the highest and the lowest BMDL (BMDL$_{highest}$/BMDL$_{lowest}$) obtained using the quantal models in Table 1. BMDL$_{highest}$/BMDL$_{lowest}$ is presented for different levels of extra risk (BMRs) and confidence levels (CLs) covered by the BMDL (a 95% confidence level is most commonly used). Incidence data on cleft palate observed in mice fetuses following maternal exposure to polychlorinated and polybrominated compounds are used as a basis. In each data-set, BMDL$_{highest}$ and BMDL$_{lowest}$ were identified considering only the models that satisfied global goodness-of-fit criteria ($p > 0.05$), as well as residual specific goodness-of-fit requirements ($\chi^2$ residual $\leq 2$ for all individual dose-groups) (Sand et al., 2002).

Hill function is illustrated in Figure 4. Each of the parameters defining this function can be interpreted easily; $\alpha$ describes the background response; $\theta$ describes the dynamic range of response (the difference between the maximum and minimum response level); $\kappa$ is the location parameter, which equals the ED$_{50}$; and $\eta$ describes the shape of the dose-response curve ($\eta$ is usually called the Hill coefficient). A similar interpretation can be used for the parameters in the exponential model, except that the location parameter, $\kappa$, does not equal the ED$_{50}$. For a value of the Hill coefficient, $\eta = 1$ the Hill function coincides with the Michaelis-Menten equation suggested to describe enzyme action (Cornish-Bowden, 1995). Considering the log-dose scale the Hill function is always S-shaped and has symmetrical properties (Figure 4). However, for $\eta \leq 1$ the Hill function cannot produce S-shaped curves on a normal dose scale; it has initially the highest slope and successively levels off to some limiting response value as the dose increases (Figure 4).

For continuous data, in addition to selecting a model describing the relationship between the dose and the mean response, a model for the variance may also be specified. A number of variance models are given in Table 2. For simplicity it is convenient to assume that the variance is constant over dose. However, for biological data this assumption is frequently not appropriate; the variance typically increases as the mean response increases, and vice versa. Due to such observations, a model for the variance may be specified as a power function of the mean response (USEPA, 2007). Of course other alternatives can be used for modeling a non-constant variance. A dose-dependent exponential function is also given in
expression for a ‘continuous benchmark response’, cBMR, is given below:

\[ cBMR = \frac{|\mu(0) - \mu(BMD)|}{\mu(0)} \]  

where \( \mu(0) \) is the background mean response and where \( \mu(BMD) \) is the response at the BMD. In the present review, cBMR is used as a general term for response definitions relating to non-probability based procedures for continuous data. The definition of response according to Equation (3) was later discussed more thoroughly and has been termed the critical effect size (CES) (Slob and Pieters, 1998). The value of the CES is intended to reflect some level of change in a toxicological endpoint that is considered acceptable on the level of the individual organism (Slob, 2002). Thus, the CES should ideally be endpoint-specific, but a consensus regarding critical effect sizes for common toxicological effect parameters may not be reached easily (Dekkers et al., 2001). More recently, this question has been discussed from a more statistical standpoint and use of the within-animal variation as a basis for establishing CESs has been suggested (Dekkers et al., 2006).

The BMD for continuous endpoints has also been defined as corresponding to a change in the mean equal to 1 control standard deviation (cBMR = 1 according to Equation (4)), in the absence of any idea of what level is adverse.

Investigations of the BMD approach for developmental toxicity has also focused on the case with continuous dose-response data (Allen et al., 1994a; Kavlock et al., 1995). In these studies, a continuous response variable was defined as the proportion of responding fetuses in each litter (Allen et al., 1994a). This response was analysed using a cBMR defined according to Equation (3). A BMDL corresponding to a cBMR of 5% was, on average, closest to the NOAEL. Note that the ‘continuous’ response definition just discussed is quite special since it is restricted to only assume values in the range between 0 and 1; the dose-response analysed concerned how the mean of the proportion of responding fetuses per litter changed with increasing dose. Both definitions of the cBMR discussed so far, i.e. Equations (3) and (4), were later applied for BMD analysis of fetal weight changes (Kavlock et al., 1995). In agreement with previous observations, a BMDL corresponding to a cBMR of 5%, according to definition (3), most resembled the

Figure 4. An illustration of the Hill function where the background, \( \alpha \), equals 10 for all cases. The dynamic range of response, \( \theta \), equals 50 for curves 1 and 2, and it equals 30 for curve 3. The location parameter, \( \kappa \), equals 10 for curves 1 and 2, and it equals 20 for curve 3. The Hill coefficient, \( \eta \), equals 1 for curve 1, while it equals 4 for curves 2 and 3. The Hill function has symmetrical properties on the log-dose scale. This figure is available in colour online at www.interscience.wiley.com/journal/jat

Table 2. From a practical point of view, the exponential function can have an advantage over the power function since the parameters for the mean and the variance becomes highly connected in the latter case which may result in estimation problems. Log-transformation of data (i.e. assume a log-normal distribution for the data) can often help to obtain a more constant variance structure, circumventing the need to use a non-constant model.

BMD Definitions for Continuous Data and Their Applications

A number of suggestions has been made for how to define the BMD for continuous endpoints. In the present review these BMD definitions are divided into two categories according to the resulting interpretation of the BMD; if it has a non-probability based or a probability based interpretation.

Non-Probability Based Interpretation of the BMD. In the early stages the BMD for continuous data was suggested as the dose causing a percentage change in response relative to the background value (Crump, 1984). This expression for a ‘continuous benchmark response’, cBMR, is given below:

\[ cBMR = \frac{|\mu(0) - \mu(BMD)|}{\mu(0)} \]  

where \( \mu(0) \) is the background mean response and where \( \mu(BMD) \) is the response at the BMD. In the present review, cBMR is used as a general term for response definitions relating to non-probability based procedures for continuous data. The definition of response according to Equation (3) was later discussed more thoroughly and has been termed the critical effect size (CES) (Slob and Pieters, 1998). The value of the CES is intended to reflect some level of change in a toxicological endpoint that is considered acceptable on the level of the individual organism (Slob, 2002). Thus, the CES should ideally be endpoint-specific, but a consensus regarding critical effect sizes for common toxicological effect parameters may not be reached easily (Dekkers et al., 2001). More recently, this question has been discussed from a more statistical standpoint and use of the within-animal variation as a basis for establishing CESs has been suggested (Dekkers et al., 2006).

The BMD for continuous endpoints has also been defined as corresponding to a change in response relative to the standard deviation in the control group, \( \sigma(0) \) (Crump, 1984, 1995; Kavlock et al., 1995):

\[ cBMR = \frac{|\mu(0) - \mu(BMD)|}{\sigma(0)} \]  

In their draft benchmark dose technical guidance document (USEPA, 2000), the U.S. EPA discusses that the BMD can be defined as corresponding to a change in the mean equal to 1 control standard deviation (cBMR = 1 according to Equation (4)), in the absence of any idea of what level is adverse.

Investigations of the BMD approach for developmental toxicity has also focused on the case with continuous dose-response data (Allen et al., 1994a; Kavlock et al., 1995). In these studies, a continuous response variable was defined as the proportion of responding fetuses in each litter (Allen et al., 1994a). This response was analysed using a cBMR defined according to Equation (3). A BMDL corresponding to a cBMR of 5% was, on average, closest to the NOAEL. Note that the ‘continuous’ response definition just discussed is quite special since it is restricted to only assume values in the range between 0 and 1; the dose-response analysed concerned how the mean of the proportion of responding fetuses per litter changed with increasing dose. Both definitions of the cBMR discussed so far, i.e. Equations (3) and (4), were later applied for BMD analysis of fetal weight changes (Kavlock et al., 1995). In agreement with previous observations, a BMDL corresponding to a cBMR of 5%, according to definition (3), most resembled the

Figure 4. An illustration of the Hill function where the background, \( \alpha \), equals 10 for all cases. The dynamic range of response, \( \theta \), equals 50 for curves 1 and 2, and it equals 30 for curve 3. The location parameter, \( \kappa \), equals 10 for curves 1 and 2, and it equals 20 for curve 3. The Hill coefficient, \( \eta \), equals 1 for curve 1, while it equals 4 for curves 2 and 3. The Hill function has symmetrical properties on the log-dose scale. This figure is available in colour online at www.interscience.wiley.com/journal/jat

Table 2. From a practical point of view, the exponential function can have an advantage over the power function since the parameters for the mean and the variance becomes highly connected in the latter case which may result in estimation problems. Log-transformation of data (i.e. assume a log-normal distribution for the data) can often help to obtain a more constant variance structure, circumventing the need to use a non-constant model.

BMD Definitions for Continuous Data and Their Applications

A number of suggestions has been made for how to define the BMD for continuous endpoints. In the present review these BMD definitions are divided into two categories according to the resulting interpretation of the BMD; if it has a non-probability based or a probability based interpretation.

Non-Probability Based Interpretation of the BMD. In the early stages the BMD for continuous data was suggested as the dose causing a percentage change in response relative to the background value (Crump, 1984). This expression for a ‘continuous benchmark response’, cBMR, is given below:

\[ cBMR = \frac{|\mu(0) - \mu(BMD)|}{\mu(0)} \]  

where \( \mu(0) \) is the background mean response and where \( \mu(BMD) \) is the response at the BMD. In the present review, cBMR is used as a general term for response definitions relating to non-probability based procedures for continuous data. The definition of response according to Equation (3) was later discussed more thoroughly and has been termed the critical effect size (CES) (Slob and Pieters, 1998). The value of the CES is intended to reflect some level of change in a toxicological endpoint that is considered acceptable on the level of the individual organism (Slob, 2002). Thus, the CES should ideally be endpoint-specific, but a consensus regarding critical effect sizes for common toxicological effect parameters may not be reached easily (Dekkers et al., 2001). More recently, this question has been discussed from a more statistical standpoint and use of the within-animal variation as a basis for establishing CESs has been suggested (Dekkers et al., 2006).

The BMD for continuous endpoints has also been defined as corresponding to a change in response relative to the standard deviation in the control group, \( \sigma(0) \) (Crump, 1984, 1995; Kavlock et al., 1995):

\[ cBMR = \frac{|\mu(0) - \mu(BMD)|}{\sigma(0)} \]  

In their draft benchmark dose technical guidance document (USEPA, 2000), the U.S. EPA discusses that the BMD can be defined as corresponding to a change in the mean equal to 1 control standard deviation (cBMR = 1 according to Equation (4)), in the absence of any idea of what level is adverse.

Investigations of the BMD approach for developmental toxicity has also focused on the case with continuous dose-response data (Allen et al., 1994a; Kavlock et al., 1995). In these studies, a continuous response variable was defined as the proportion of responding fetuses in each litter (Allen et al., 1994a). This response was analysed using a cBMR defined according to Equation (3). A BMDL corresponding to a cBMR of 5% was, on average, closest to the NOAEL. Note that the ‘continuous’ response definition just discussed is quite special since it is restricted to only assume values in the range between 0 and 1; the dose-response analysed concerned how the mean of the proportion of responding fetuses per litter changed with increasing dose. Both definitions of the cBMR discussed so far, i.e. Equations (3) and (4), were later applied for BMD analysis of fetal weight changes (Kavlock et al., 1995). In agreement with previous observations, a BMDL corresponding to a cBMR of 5%, according to definition (3), most resembled the
NOAEL. Considering definition (4) the same was the case for a cBMR of 0.5.

Similar to the situation for quantal data, the observations discussed above have been used as a rationale for specifying values of the cBMR in other studies (Allen et al., 1996, 1998). The BMD suggested for boric acid (using fetal weight data) in Allen et al. (1996) was based, for example, on a 5% change in response relative to background. In their cumulative risk assessment of organophosphates the U.S. EPA used the dose causing a 10% inhibition of brain cholinesterase as a starting point (Reiss and Gaylor, 2005). The definition of the BMD as the dose causing a 5% and 10% change in response relative to background has also been used in applications to developmental neurotoxicity, using 2,2',4,4',5-pentabromodiphenyl ether (PBDE99) as a model substance (Sand et al., 2004).

Besides the definitions discussed above, the BMD has also been presented as the dose causing a certain change in response relative to the dynamic range of response (the difference between the maximum and the minimum response level) (Murrell et al., 1998):

\[ cBMR = \frac{\left| \mu(0) - \mu(BMD) \right|}{\text{dynamic range of response}} \]  

This way of expressing response is quite commonly adopted when models that include a parameter describing the dynamic range of response are used (e.g. Hill function and exponential model, Table 2). The U.S. EPA has also suggested that the observed rather than the estimated dynamic range may be used when other models are considered (USEPA, 2000). Continuous responses may differ in terms of the dynamic range of response, and Murrell et al. (1998) argued that a certain change under Equation (5) may be comparable across different endpoints (which may not be the case for the CES, Equation (3)). Murrell et al. (1998) also pointed out that their suggested response definition is effectively identical to the extra risk definition for quantal data, i.e. Equation (2). The Murrell et al. (1998) response definition has been employed, for example, when calculating BMDs for TCDD based on biochemical responses (Kim et al., 2002).

Recently, Sand et al. (2006) discussed that the BMD may be defined as the dose where the slope of the S-shaped dose-response curve changes the most per unit log-dose. A general model known as the Richards function was assumed for the S-shaped dose-response. As shown, log BMD\(_T\) corresponds to the dose where the second derivative of the dose-response function (Richards function) is the lowest, and it can be solved as one of the roots of the second derivative. In the example, the response at log BMD\(_T\) under the Murrell et al. (1998) definition (Equation 5) is \(\approx 0.15\). In Sand et al. (2006) two special cases of the Richards function were studied in detail; the Hill function (symmetrical case) and the Gompertz curve (the most asymmetrical case). For the Hill function and the Gompertz curve the corresponding responses at log BMD\(_T\) are 0.21 and 0.073, respectively (the response at log BMD\(_T\) decreased with increasing asymmetrical properties of the curve). As a conservative approach, a response of 5–10% was suggested for BMD estimation; this is close to the ‘change point’ in the case of Gompertz characteristics, or below the ‘change point’ for curves which are less asymmetrical and Gompertz-like. This figure is available in colour online at www.interscience.wiley.com/journal/jat
Probability Based Interpretation of the BMD. There are also definitions of the BMD for continuous endpoints that focus on making statements in terms of probability (or risk) in similarity to the case for quantal data, attempting to generalize the BMD approach. A probability based interpretation of the BMD for continuous data can be obtained by the application of a cut-off value denoting an ‘adverse’ response. A cut-off point may represent some theoretical response level that is considered to be associated with ‘adverse’ health effects. By categorizing experimental observations according to a cut-off, the continuous response variable becomes dichotomized. The transformed data can then be analysed using the common quantal dose-response models (Table 1). This type of methodology has, however, been criticized due to the fact that information is lost in the process of data transformation (Allen et al., 1994a; Gaylor, 1996; West and Kodell, 1999; Crump, 2002).

An alternative procedure was presented by Gaylor and Slikker (1990), which sometimes is referred to as the ‘hybrid approach’ (Crump, 2002); a terminology that will be adopted in the present review. The hybrid approach is illustrated in Figure 6. An important aspect of this method involves estimating the distribution of the data. As is shown in Figure 6, a probability model can be established that describes the proportion of the distribution that is below (or above) a cut-off point as a function of dose. For normally distributed data, with constant variance over dose, the equation for the probability model, \( p(d) \), equals

\[
p(d) = \Phi\left[\frac{c - \mu(d)}{\sigma}\right],
\]

for a response that decreases with increasing dose of chemical, where \( \Phi \) is the cumulative standard normal distribution function; \( c \) is the cut-off value; \( \mu(d) \) is the mean response at dose \( d \); and where \( \sigma \) is the standard deviation. The BMD can be expressed as the dose where the probability of falling below the cut-off level has increased by some value according to the additional or extra risk definition.

A critical aspect of decision making in the hybrid procedure is the determination of the cut-off value, \( c \). Generally, determination of a cut-off point is performed in a statistical sense. For example, adverse or extreme observations may be defined as responses 2–3 standard deviations from the control mean. Similarly, the cut-off can be defined as corresponding to an extreme tail proportion of the control distribution; corresponding to a certain percentile. For example, the cut-off, \( c \), in Equation (6) may be solved as corresponding to some specified value of \( p(0) \). Considering experimental data, the cut-off level has been suggested to correspond to a \( p(0) \) in range of 0.01–0.05 (Crump, 1995; Kodell et al., 1995). For epidemiological studies, \( p(0) = 0.05 \) has been suggested with the argument that it corresponds to the definition of the normal range for clinical data (Crump et al., 1998, 2000).

The hybrid methodology was initially illustrated for neurochemical and neurohistological effect data observed in monkeys and rats exposed to methylenedioxyamphetamine (Gaylor and Slikker, 1990, 1994). Further application of the approach to experimental neurotoxicity data has also been presented (Kodell et al., 1995; Slikker et al., 1998). In addition, the hybrid model has been applied to neurobehavioral endpoints (Sand et al., 2004; Zhu et al., 2005). Besides applications to neurotoxicity data, the hybrid approach has been used to calculate BMDs for systemic effects such as changes in relative liver weights observed in mice after exposure to trichloroethylene (Barton and Das, 1996), and changes in body and lung weights observed in rats in connection with exposure to nickel compounds (Haber et al., 1998). By defining the cut-off as corresponding to the 1st or 99th percentile of the control distribution (\( p(0) = 0.01 \)), the hybrid approach has been compared with a case where the BMD was defined as corresponding to a 1% change in response under the Murrell et al. (1998) response definition (Equation (5)) (Gaylor and Aylward, 2004).

Figure 6. An illustration of the hybrid approach considering a hypothetical continuous response variable. The mean response as a function of dose is described by the power model \((\alpha = 12, \beta = -0.4, \eta = 1.4)\). The distribution of the data is assumed to be normal with standard deviation \( \sigma = 1.0 \). A cut-off value denoting ‘adverse response’ is defined as corresponding to the 5th percentile of the control distribution, \( p(0) = 0.05 \), which equals a continuous response of 10.36. The probability model, \( p(d) \), describes the proportion of the distribution that is below the cut-off as a function of dose, i.e. Equation (6). The BMD may be defined as corresponding to a 5–10% increase in the probability of falling below the cut-off according to the additional or extra risk definition, i.e. Equation (1) or (2). This figure is available in colour online at www.interscience.wiley.com/journal/jat
This latter procedure was called the ED$_{01}$ approach by the authors. Gaylord and Aylward (2004) used the Hill function in their analysis and concluded that the uncertainty associated with BMDs corresponding to additional risks of 1%, 5% and 10% (estimated under the hybrid approach) were lower than that associated with the ED$_{01}$. These conclusions were attributed to the fact that a higher degree of extrapolation was associated with the ED$_{01}$ approach, i.e. the ED$_{01}$ was generally much lower than the BMDs estimated under the hybrid method.

The hybrid concept has to some extent been applied to human epidemiological data. For epidemiological data, the methodology is applicable to studies where the response and exposure have been recorded for each individual. Several epidemiological studies applying the hybrid methodology have focused on neurological effects in association with mercury/methylmercury exposure (Budtz-Jorgensen, 2000; Crump et al., 1995, 1998, 2000), and also in association with exposure to polychlorinated biphenyls (Jacobson et al., 2002) and manganese (Clewell et al., 2003). The U.S. EPA derived a reference dose for methylmercury in 2001 that was based on the hybrid approach. In the calculations they selected a $p(0) = 0.05$, and a BMR = 0.05 (Rice, 2004). The renal effects of cadmium have also been assessed using the hybrid procedure (Suwazono et al., 2006).

Methodological aspects of the hybrid approach have been discussed (Stiteler and Swartout, 1991; Kodell and West, 1993; West and Kodell, 1993, 1999; Crump, 1995, 2002; Gaylord and Chen, 1996; Sand et al., 2003, 2004). For example, Crump (1995) discussed the equivalence between the hybrid approach and the procedure where the BMD is defined as corresponding to a change in response relative to the standard deviation in the control group, i.e. Equation (4). The U.S. EPA suggestion that the BMD can be defined as corresponding to a change in the mean equal to 1 control standard deviation indirectly suggests use of the hybrid approach with $p(0) = 0.02$, and BMR = 0.10 (USEPA, 2000). In Sand et al. (2003) the influence of variance on the hybrid model was studied. This analysis showed that if the cut-off point is defined as corresponding to a percentile of the control distribution, the BMD becomes biased upward when the variance is biased upward. Conversely, if the cut-off is defined independently of the model estimated from the sample data (i.e. directly as some level of the continuous response variable), the BMD becomes biased upward when the variance is biased downward (Sand et al., 2003).

**Critical Issues in BMD Analysis**

**Definition of the BMD: Summary**

A major aim of the present review has been to present the different procedures that have been used and suggested for BMD calculations. A BMR (additional or extra risk) of 5% or 10% has generally been applied in calculations for quantal data. This has been based on the resemblance between the resulting BMDL and the NOAEL, on average (Allen et al., 1994a, 1994b; Fowles et al., 1999; Haag-Grönlund et al., 1995), and it has also been supported (relative to the 1% level) considering the impact of model dependence at lower response levels (Sand et al., 2002, Figure 3). Based on these findings risk levels of 5% or 10% seem reasonable as defaults.

For the case of continuous endpoints, several BMD definitions have been proposed in the scientific literature (Crump, 1984; Gaylord and Slikker, 1990; Murrell et al., 1998; Slob and Pieters, 1998; Sand et al., 2006). A short summary and some important remarks regarding the procedures discussed are given below:

The critical effect size (CES) approach (Slob and Pieters, 1998); the definition of the BMD as corresponding to a percentage change in response relative to background is a reasonable definition and from a practical point of view it can be used quite generally (i.e. it does not require that information of the maximum/minimum response is available). However, this approach ideally requires endpoint specific response values, which may not be easily established. A CES of 5% is typically suggested as default for experimental data (in the absence of endpoint specific suggestions) (Woutersen et al., 2001). Certain studies have shown that a 5% response results in a BMDL that is close to the NOAEL, on average (Allen et al., 1994a; Kavlock et al., 1995).

The definition of the BMD as corresponding to a change relative to the dynamic range of response (Murrell et al., 1998) provides a means of standardization for continuous data. A given change under this response definition is probably more comparable across different types of endpoints compared with a certain CES. This response definition ideally requires that information of the maximum/minimum response value is available which may not always be the case. An extension of this approach, which focuses on defining the BMD where the slope of the S-shaped dose-response changes the most per unit log-dose (Sand et al., 2006, Figure 5), introduces a ‘change point’ in the low dose-region which may be relevant from a risk assessment point of view. Responses of 5–10% have been applied and suggested under the Murrell et al. (1998) definition (Kim et al., 2002; Sand et al., 2006).

The hybrid concept (Gaylord and Slikker, 1990; Crump, 1995, Figure 6) introduces the possibility of making statements in terms of probability (or risk) in similarity to the case for quantal data. In the hybrid approach an additional risk of 5–10% (as in the case for quantal data) is commonly used and the cut-off, $p(0)$, is generally specified to 5%. This procedure has been questioned in the case of experimental data. The standard deviation is an important part of this model and its estimate can be

*Copyright © 2007 John Wiley & Sons, Ltd.*
Selection of Dose-Response Model and BMDL

The selection of the dose-response model for BMD and BMDL estimation clearly represents one of the important issues associated with the BMD concept. From a toxicological point of view many continuous endpoints can be expected to have an S-shaped dose-response relationship. For estimation of S-shaped models (such as the Hill function), the underlying data set must contain doses both in the low, middle and high dose range. Since typical sigmoidal models for continuous data have four free parameters (e.g., Hill function and exponential model; Table 2), at least five dose groups are required to prevent over-parameterization and to enable a statistical evaluation of the model fit. Even though it may be desirable to apply these models for continuous endpoints the necessary data requirements may vary on a case-by-case basis. For example, the estimation of the dynamic range of response (or maximum response) can be problematic if data in the high dose region are lacking. Thus simpler models, or special cases of the general S-shaped model used, often need to be considered. In the case of quantal data, this is not a problem to the same extent since the quantal response is bounded between 0 and 100%; a specific parameter describing the dynamic range of response is not required since all the typical models (Table 1) level off at a response of 100%.

When selecting a dose-response model for BMD analysis an option is to consider several model classes, or only a certain model class. In the case of quantal data it is common to consider all the different model classes in Table 1, while for continuous data it has been more typical to focus on a specific model family in the BMD analysis. Even though a single model class is considered, different special cases of that class may be tested; for example, several special cases can be derived from the polynomial model, the Hill function and the exponential model family in Table 2, respectively. In any case, dose-response models in the BMD field are used descriptively and a basic strategy is to define a criterion for adequate model fit. The U.S. EPA suggest that models should be discharged if the overall p value is ≤ 0.10 (USEPA, 2000), while a more traditional cut-off is p ≤ 0.05. Further model selection can, however, usually be necessary since several models may meet the default criterion for adequate fit.

In cases where models from the same class were only considered, the use of the likelihood ratio test can be used to address the issue of further model selection. This is a fairly straightforward approach by which it can be determined whether or not two models (belonging to the same class) fit the data significantly differently. If not, the model with a fewer number of free parameters is selected; the general rule in model selection is to only consider the more complex model if it fits the data significantly better than the simpler model (with fewer parameters).

In the case where a pool of several model classes is used, other consideration needs to be made for further model selection. Different suggestions have been made in this respect. For example, the Akaike’s Information Criterion (AIC) has been suggested (USEPA, 2000). According to this approach the model with the lowest AIC value is selected; the AIC is a measure of the degree of fit weighted by the number of free parameters in the model. In contrast to the likelihood ratio test grading under the AIC does not involve making statements whether or not a certain model fits the data significantly better than another model. Thus, in cases when AIC values between the models are rather similar (for a certain data set), model ranking according to the AIC may not be very powerful (Burnham and Anderson, 1998; Sand et al., 2002). A conservative model selection approach appears more straightforward in the case of different model classes; i.e. to select the lowest BMDL considering models that adequately fit the data. As an extension, if several special cases of each model class are considered, the most appropriate member (which adequately fits the data) within each class can first be determined, based on likelihood ratio tests, and then the lowest BMDL can be selected.

The option to perform some form of averaging of BMD (or BMDL) values instead of using the result from a single model has also been discussed to some extent (Fitzgerald et al., 2004; Bailer et al., 2005). For example, Bailer et al. (2005) suggests the use of Bayesian model averaging. According to this approach a collection of models (e.g. those in Table 1) is selected in advance and averaging is made using a weighted method where the contribution from a particular model depends on the support the data suggest for that model.

It should also be pointed out that while it is helpful to have a formal process, as described above, according to which decisions can be made, one should not forget that visual examination of the plots is also very informative. For example, visual analysis allows the investigator to confirm that the results from testing (which are only numbers) are reasonable; it helps to identify possible outliers that may have an impact on the results; and it helps to conclude whether or not extrapolation is
involved in the BMD analysis. Visual evaluation of the fitted curve also helps to confirm that it does not have an un-plausible form; if constraints are not imposed on the model parameter spaces prior to fitting the model, estimated parameters could potentially assume unreasonable values, e.g. resulting in a ‘wavy’ curve. In a more general sense, investigating plots is important since it contributes to make the risk assessor more familiar and experienced with the BMD concept as such.

Further Issues in BMD Analysis

Study Design

An important aspect of the BMD approach concerns whether or not current experimental study designs, typically including a control and three treatment groups (e.g. as recommended in the OECD guidelines in sub-chronic toxicity studies), are appropriate for BMD and BMDL estimation. For quantal data analysis of this issue has focused on developmental toxicity and the estimation of a BMD corresponding to an excess risk of 5% (Weller et al., 1995; Kavlock et al., 1996; Krewski et al., 2002). Kavlock et al. (1996) concluded that the standard designs are adequate for BMD calculations but may be improved by minor modifications. Krewski et al. (2002), argued that sub-optimal designs including more than three dose groups are preferable since construction of the near-optimal design (a design with three to four dose groups including control was close to optimal) requires a priori knowledge of the underlying dose-response curve (Krewski et al., 2002). That the use of several dose groups could be more beneficial from a practical point of view has also been indicated by Weller et al. (1995).

The issue of study design in BMD calculations has also been discussed for continuous endpoints (Woutersen et al., 2001; Slob et al., 2005; Kuljus et al., 2006). Slob et al. (2005) considered the case of defining the BMD as corresponding to a CES of 5%. They concluded that the performance of a design is determined by the total number of animals used, and distributing them over more dose groups does not result in a poorer performance of the study. In Kuljus et al. (2006) the case of defining the BMD as corresponding to a 5% change in response relative to the dynamic range of response was considered. Based on minimizing the variance of the BMD as design criterion and assuming that the dose-response follows the Hill function, Kuljus et al. (2006) concluded that it is preferable to distribute animals over more than four dose groups (including the control). Using four dose groups for estimating the S-shaped curve is not straightforward (e.g. it requires that doses are placed optimally); it represents a case of over-parameterization indicating that use of a simpler dose-response model is probably to be preferred in such a case.

Taken together, these studies indicate that while the standard designs are adequate for BMD analysis, it appears generally to be beneficial to distribute animals over more dose levels, e.g. since the optimal placement of doses (for dose-response modeling) in the given situation is not known beforehand and since the use of several dose groups could minimize the impact of any systematic errors that may be present in individual dose-groups.

Comparison of Dose-Response Curves and BMDs

Besides providing a possibility to obtain improved points of departures for risk assessment, the use of the BMD concept and dose-response modeling, in a more general sense, also gives opportunities to compare dose-response characteristics between different sub-populations, animal strains or species, and chemicals. By the introduction of covariates in the analysis it is for example possible to investigate what aspects of the dose-response relationship are similar, or different, in two sub-populations. The use of a BMD ratio is a natural approach for estimating the difference in sensitivity between two species, or the relative potency between two chemicals.

The use of covariates in dose-response modeling has generally been discussed by Slob (2002). In Sand et al. (2004), such a type of analysis was applied to investigate possible differences between male and female rats in their neurobehavioral response to $2,2',4,4',5$-pentabromodiphenyl ether (PBDE99). Bokkers and Slob (2005) have discussed an approach to derive a data-based assessment factor for sub-chronic to chronic extrapolation by the calculation of BMD ratios. In this study, a comparison was also made using NOAEL ratios for the same purpose. It was concluded that the BMD ratio distribution was less wide compared with the NOAEL ratio distribution, indicating that a higher precision and less conservatism is obtained when a dose-response modeling approach is applied (Bokkers and Slob, 2005). The use of NOAEL ratios, in comparative analysis, can also be criticized in a more general sense since two NOAELs are not really comparable quantities. On the contrary, two BMDs are comparable if they have the same definition; i.e. the dose causing a given response (BMR or cBMR). Since the BMD has a specific definition that relates to the potency of the chemical under a given condition (in a certain species, for a certain route of exposure, etc), the BMD concept is better suited for comparative analysis relative to the NOAEL approach.

The issue of comparing dose-response curves and BMDs has also been investigated by Sand et al. (in manuscript) considering dioxin sensitive Long-Evans (L-E) rats and dioxin resistant Han/Wistar (H/W) rats following long-term exposure to TCDD. The difference in sensitivity between L-E and H/W rats for a given endpoint was quantified in terms of a BMD ratio (i.e.
BMD\textsubscript{H/W}/BMD\textsubscript{L/E}). It is important to note that a BMD ratio is dependent on the level of response (BMR, or eBMR) selected if the dose-response curves fundamentally differ in terms of their shapes. However, considering the case of ‘parallel’ dose-response curves, which have similar shapes, the BMD ratio is constant and reflects the difference in the location of the curves on the dose scale\textsuperscript{1}. In Sand et al. (in manuscript) such and related questions were addressed using a likelihood ratio test statistic; the concept of parallel curves was tested, it was investigated if the strains differed significantly in their response to TCDD treatment (i.e. it was tested if BMD\textsubscript{H/W}/BMD\textsubscript{L/E} = 1, in which case the L-E and H/W dose-response curves have identical location), and a confidence interval for the BMD ratio was established to describe the uncertainty in the estimated sensitivity difference\textsuperscript{3}.

The related issue of comparing dose-responses for analysing relative potency differences has been discussed by Toyoshiba et al. (2004) for the case of a dioxin-like compound. In this analysis the assumption of parallel curves for liver enzyme activity could not be supported, according to likelihood ratio test methods, indicating that for some endpoints a single relative potency factor is not appropriate for the comparison of dose-response behavior of different dioxin-like congeners.

**Summary and Conclusion**

This review has described the current state of knowledge of the BMD concept for health risk assessment of chemicals. While the BMD method has been applied in the USA it has not been considered to the same extent in Europe. There is, however, a general trend of expanding focus on statistical method developments in the area of quantitative health risk assessment within the European Community, and discussion and awareness of the BMD concept has increased.

\textsuperscript{1} A BMD ratio that is constant and not dependent on the value of BMR, or eBMR, results if certain constraints are imposed when fitting the two dose-response curves. The constraints required for this may be different depending on what BMD definition is used. In Sand et al. (in manuscript) the definition of the BMD as corresponding to a \% change in response relative to background (Slob and Pieters, 1998), and the definition of the BMD as corresponding to a \% change in response relative to the dynamic range of response (Murrell et al., 1998) were used. ‘Parallel curves’ herein refers to a certain constraint under which the BMD ratio is constant under both these BMD definitions.

\textsuperscript{2} The likelihood ratio test can be used for selecting the most appropriate member within a certain model class, and for estimating the confidence interval for the BMD (and hence the BMDL). This test can also be used to investigate if dose-response curves are parallel, if two chemicals have different potency, or if two species differ in their sensitivity to a certain exposure. It can also be used for estimating a confidence interval for a BMD ratio, reflecting potency or sensitivity differences (similar to estimating a confidence interval for the BMD). Thus, the likelihood ratio test is a quite useful tool for addressing many of the issues that are associated with the BMD method and dose-response modeling.

A highlighted feature of the BMD concept is that it involves uncertainty analysis by the establishment of a confidence interval and use of the BMDL. This is a major improvement relative to the NOAEL approach. Other arguments in favor of the BMD relative to the NOAEL include that the BMD approach makes more use of all the available data, and that the resulting starting point for risk assessment is not restricted to being one of the experimentally selected dose levels. Studies have also indicated that the NOAEL should not be considered a risk/response-free exposure level. In fact, the response at the NOAEL may in several cases exceed those that are typically suggested to correspond to the BMD. In addition, the BMD is more suitable for comparative analysis relative to the NOAEL; for example a BMD ratio reflects a potency difference between two chemicals, or a sensitivity difference between two populations.

The definition of the BMD represents a critical issue. For quantal data the BMD is generally defined as corresponding to an additional or extra risk of 5\% or 10\%. While a benchmark response of 1\% was also suggested in the original definition of the BMD, this response level is usually not applied in practice since this can often result in extrapolation outside the experimental range. The application of the BMD concept to continuous data has merited more discussion; several BMD definitions are available which have been discussed in the present review. It is proposed that the BMD for continuous data is defined as corresponding to a percentage change in response relative to background or relative to the dynamic range of response. Under these definitions, a 5\% or 10\% change can be regarded as default. Concerning the definition of the BMDL it is usually specified as the one-sided lower 95\% confidence limit on the BMD.

Another important aspect of the BMD method involves the selection of the dose-response models to use as a basis in the establishment of the BMD and BMDL. There are a number of models suggested for quantal as well as continuous data. This implies that some comparison of how well the different models describe a particular data set needs to be performed. The general approach is to further consider models, which are not rejected under some default statistical criterion (i.e. \( p \leq 0.05 \)). The determination of a BMD and BMDL given a number of acceptable models (that were not rejected) may then (1) be based on statistical comparison between the models if they belong to the same class, or (2) be the most conservative value (the lowest BMDL) if the acceptable models belong to different classes. A combination of (1) and (2) could also be used; the most appropriate member within each model class is first determined based on statistical considerations, and then the lowest BMDL is selected.

Default approaches for BMD analysis are available which allows an application of the method in many cases. In our opinion the BMD concept clearly represents a
methodological improvement in the field of health risk assessment of chemicals. To make better use of the approach more emphasis should be placed on selecting study designs appropriate for dose-response modeling.

References


