From the Institute of Environmental Medicine,
Division of Environmental Health Risk Assessment,
Karolinska Institutet, Stockholm, Sweden

DOSE-RESPONSE MODELING

Evaluation, Application, and Development of Procedures for Benchmark Dose Analysis in Health Risk Assessment of Chemical Substances

Salomon Sand

Stockholm 2005

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 91-7140-420-1

Till mamma

ABSTRACT

In this thesis, dose-response modeling and procedures for benchmark dose (BMD) analysis in health risk assessment of chemical substances have been investigated.

The BMD method has been proposed as an alternative to the NOAEL (no-observed-adverse-effect-level) approach in health risk assessment of non-genotoxic agents. According to the BMD concept, a dose-response model is fitted to data and the BMD is defined as the dose causing a predetermined change in response. A lower statistical confidence limit on the BMD (the BMDL) has been suggested as the point of departure in the determination of guidance values such as acceptable daily intakes (ADIs). The fact that the BMD corresponds to an explicit response level has been argued as a major advantage since this introduces consistency and suggests that the point of departure for risk assessment will be based on more information. However, this feature also represents one of the challenges associated with the BMD approach, i.e. which level of response should the BMD correspond to? This and related questions have been addressed in the present thesis.

Given the original definition of the BMD as the dose causing a 1 – 10% increase in the risk for adverse heath effects compared to background, this study has analyzed the impact of model choice in BMD calculations. In the case of quantal data, it is suggested that the BMD is defined as corresponding to risk levels in the range of 5 - 10%.

The introduction of the BMD method for continuous endpoints has merited more discussion, and the present thesis has mainly focused on these issues. A probability based procedure suggested for continuous data has been analyzed in detail. By the definition of a cut-off value denoting “adverse” response, this approach allows the BMD to be interpreted as corresponding to some risk level, similar to the case for quantal data. While this may be of interest, it was shown that the choice of determination of the cut-off point is of high importance and dictates how the BMD depends on the variance.

In the present thesis, the BMD approach was introduced for neurobehavioral endpoints using 2,2’,4,4’,5-pentabromodiphenyl ether (PBDE99) as a model substance. According to proposed methods of BMD analysis for continuous data, a BMDL of about 0.5 mg/kg bw was obtained for PBDE99. In another application, statistical differences in sensitivity between dioxin sensitive Long-Evans (L-E) and dioxin resistant Han/Wistar (H/W) rats following long-term exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin were demonstrated for the first time. Differences between L-E and H/W rats were most pronounced for volume fraction of hepatic foci; L-E rats were approximately 80 times more sensitive than H/W rats. Considering data on body and organ weights, L-E rats were 10-20 times more sensitive than H/W rats. For retinoid parameters, and hepatic CYP1A1 induction, differences between the strains were generally about 5-fold, and associated with a low uncertainty.

Besides analysis and application of suggested procedures for BMD analysis of continuous endpoints, developments have also been proposed. For dose-response relationships that are S-shaped, the BMD may be defined as the dose where the slope of the curve changes the most in the low dose region. It is discussed whether this definition may provide a biological basis for selection of the response associated with the BMD. Considering a conservative scenario, it was shown that the dose where the slope changes the most corresponds to a response in the range of 5 - 10%, if defined as a percentage change in response relative to the magnitude of response. This definition of the BMD may be considered for continuous data in future applications.

LIST OF PUBLICATIONS

The thesis is based on the following original papers which will be referred to by their Roman numerals:

I. **Sand**, **S**., Falk Filipsson, A., Victorin, K. Evaluation of the Benchmark Dose Method for Dichotomous Data: Model Dependence and Model Selection. *Regul Toxicol Pharmacol 36*, 184-197, 2002.

II. **Sand**, **S**., von Rosen, D., Falk Filipsson, A. Benchmark Calculations in Risk Assessment Using Continuous Dose-Response Information: The Influence of Variance and the Determination of a Cut-Off Value. *Risk Anal 23*, 1059-1068, 2003.

III. **Sand**, **S**., von Rosen, D., Eriksson, P., Fredriksson, A., Viberg, H., Victorin, K., Falk Filipsson, A. Dose-Response Modeling and Benchmark Calculations from Spontaneous Behavior Data on Mice Neonatally Exposed to 2,2',4,4',5-Pentabromodiphenyl Ether. *Toxicol Sci 81*, 491-501, 2004.

IV. **Sand**, **S**., Fletcher, N., von Rosen, D., Victorin, K., Viluksela, M., Tuomisto, J.T., Tuomisto, J., Falk Filipsson, A., Håkansson, H. Quantitative and Statistical Analysis of Differences in Sensitivity Between Long-Evans and Han/Wistar Rats Following Long-Term Exposure to 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin. *Manuscript*, 2005.

V. **Sand**, **S**., von Rosen, D., Victorin, K., Falk Filipsson, A. Identification of a Critical Dose Level for Risk Assessment: Developments in Benchmark Dose Analysis of Continuous Endpoints. *Submitted*, 2005.

CONTENTS

LIST OF ABBREVIATIONS

ADI	Acceptable daily intake
AIC	Akaike's information criterion
BMD	Benchmark dose
BMDL	Lower confidence bound on the benchmark dose
BMDS	Benchmark dose software
BMDTG	Benchmark dose technical guidance document
BMR	Benchmark response
cBMR	Continuous benchmark response
CES	Critical effect size
H/W	Han/Wistar
L-E	Long-Evans
LOAEL	Lowest-observed-adverse-effect level
NOAEL	No-observed-adverse-effect level
PBDE99	2,2',4,4',5-pentabromodiphenyl ether
POD	Point of departure
TCDD	2,3,7,8-tetrachlorodibenzo-*p*-dioxin
USEPA	United States Environmental Protection Agency
WHO	World Health Organization

1 BACKGROUND

1.1 GENERAL INTRODUCTION

According to the World Health Organization (WHO), the process of health risk assessment of chemicals is divided into four steps (WHO, 1999):

- Hazard identification
- Dose-response assessment
- Exposure assessment
- Risk characterization

Hazard identification is the identification of the inherent capability of a substance to cause adverse health effects. Chemical hazards are usually identified from *in vivo* experiments in laboratory animals. Important information may also be derived from *in vitro* studies, particularly with respect to genotoxic and mutagenic effects. In some cases, chemical hazards have initially been recognized in human epidemiological studies.

Dose-response assessment involves characterizing the relationship between the dose of a chemical agent and the biological effects that are produced. This analysis is usually performed in controlled animal experiments; groups of animals are exposed to specified doses of the chemical and the response to treatment is investigated. While it may be of interest to use human epidemiological data as a basis for the dose-response analysis, this is generally more problematic since exposure information in these types of studies often is limited. The final aim of the dose-response assessment is to determine a dose level of the chemical substance that may be used as a starting point, i.e. a point of departure (POD), in the establishment of acceptable exposure levels for the human population, including sensitive subgroups (e.g. children). It is the process of dose-response assessment that has been the focus of the present thesis.

Exposure assessment involves estimating the degree to which the human population is exposed to a hazardous agent. This can be performed qualitatively (e.g. from questionnaires) and/or quantitatively. In the latter case, a distribution describing the exposure in a given population can be estimated from collected data. Such an approach focuses not only on the average exposure/intake but also on the variability in the exposure that is apparent in a population.

Finally, risk characterization is defined as the integration and joint evaluation of hazard identification, dose-response assessment, and exposure assessment. For example, by comparing the exposure situation with outputs from the dose-response analysis (i.e. the POD) information of the margin of safety can be obtained. It may thus be concluded whether a population, or a certain subgroup, potentially is at risk. Another purpose of the risk characterization is to identify exposure levels, like acceptable daily intakes (ADIs), which are considered to be safe for the human population.

1.1.1 Non-genotoxic effects; threshold assumption

For the majority of toxicological effects, excluding genotoxic effects, it is assumed that there exists an exposure threshold below which there are no biologically significant outcomes. 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). However, it is problematic to experimentally prove the presence or absence of such a threshold; an experiment may include doses without a measurable (statistically significant) biological effect whether or not the dose-response relationship has a threshold (Edler *et al*., 2002). Also, from a practical point of view it can be difficult to perform studies involving very low exposure levels.

In the scope of the threshold assumption, the traditional approach in health risk assessment of non-genotoxic agents 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 to the response in the control group. In Figure 1, the NOAEL and the LOAEL (lowest-observed-adverse-effect-level) are illustrated for malformation dose-response data observed in mice fetuses following maternal exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). The NOAEL is traditionally used as a POD in the determination of guidance values like ADIs.

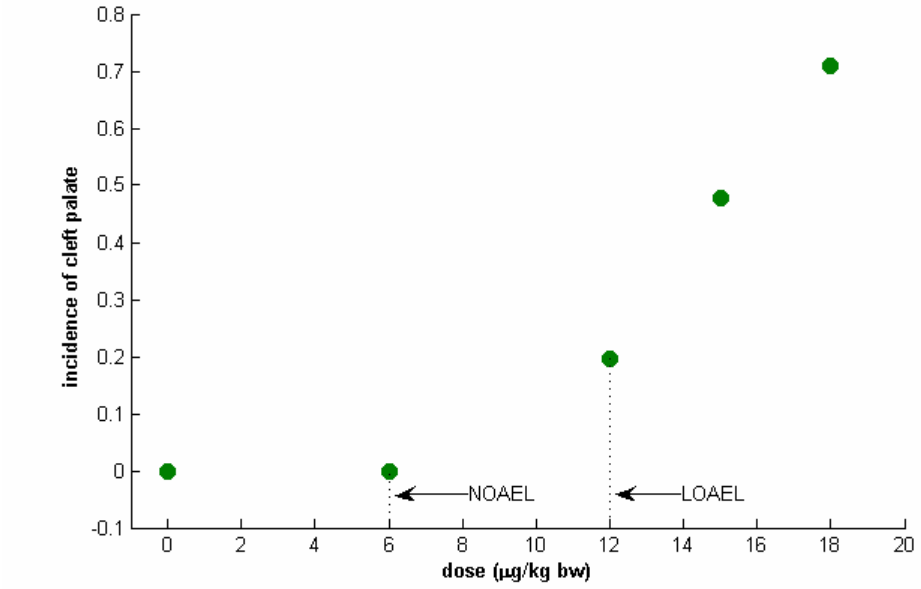

Figure 1. Incidence of cleft palate observed in mice fetuses following maternal exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) (data from Birnbaum *et al*., 1989). The no-observed-adverse-effect level (NOAEL) and the lowest-observed-adverse-effect level (LOAEL) are 6 and 12 µg/kg bw, respectively.

The standard procedure for establishment of guidance values such as ADIs is to apply uncertainty factors to the NOAEL. For experimental data, an uncertainty factor is used for animal to man extrapolation. Generally, a factor is also used to account for differences in sensitivity in the human population. These two factors are each by tradition assigned a value of 10 (WHO, 1999). Thus, to obtain a guidance value the NOAEL is by default divided by a factor 100 (10 × 10). Additional uncertainty factors may also be used depending on characteristics of the study for which the NOAEL was derived, for example taking into account the duration and route of exposure.

1.1.2 Genotoxic carcinogens; non-threshold assumption

For genotoxic carcinogens it is considered that even very low levels of exposure may increase the risk for adverse outcomes, i.e. there exits no exposure threshold below which the risk for adverse effects is zero. This assumption is based on the idea that a single molecule of genotoxicant may be sufficient to cause a DNA damage that eventually may result in the development of a tumor (USEPA, 1995). Because of this, for genotoxic carcinogens, cancer risk is traditionally assumed to be proportional to dose at low doses. While such an assumption is difficult to prove, there are examples where it has been supported experimentally. For instance, in a study by Peto *et al*., (1991) involving a large number of rats exposed to nitrosamine (*N*-nitrosodiethylamine and *N*-nitrosodimethylamine) a linear relationship was observed between dose and the incidence of liver neoplasms, at doses below 1 ppm (volume in drinking water).

Several dose-response models have been proposed 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, i.e. 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 to be 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 carcinogen risk assessment, 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, and 10%. Any of theses values may then be considered as a POD, typically subject to linear extrapolation (USEPA, 2005a).

1.2 THE BENCHMARK DOSE METHOD

The use of a NOAEL in setting standards for human exposure to non-genotoxic agents has been criticized (Crump, 1984; Kimmel and Gaylor, 1988; Leisenring and Ryan, 1992; Allen *et al*., 1994a; Barnes *et al*., 1995; USEPA, 1995). Shortcomings associated with this procedure for health risk assessment include the following;

- The NOAEL is limited to the doses tested experimentally.
- The shape of the dose-response relationship is not considered in the determination of the NOAEL.
- Experiments involving fewer animals tend to produce larger NOAELs. The reverse would seem more appropriate, arguing that larger experiments should provide greater evidence of safety. The rationale for this conclusion is that when a large number of animals are used it is more likely that a statistical difference between two groups may be detected.
- Use of a NOAEL does not provide an estimate of the potential risk, or effect, associated with the exposure of interest.

Due to the limitations identified with the NOAEL concept, the benchmark dose (BMD) method has been suggested as an alternative approach to be used in health risk assessment. The present thesis concentrates on investigating the BMD concept, which 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. The lower 95% confidence limit on the BMD, i.e. the BMDL, has been proposed to replace the NOAEL as a POD for determination of guidance values, i.e. uncertainty factors are applied to the BMDL instead of to the NOAEL. The BMD method is illustrated in Figure 2, using malformation data observed in mice fetuses following maternal exposure to TCDD as an example. A number of advantages of the BMD methodology compared to the traditional NOAEL procedure have been identified (Crump, 1984; Kimmel and Gaylor, 1988; Leisenring and Ryan, 1992; Allen *et al*., 1994a; Barnes *et al*., 1995; USEPA, 1995; Edler *et al*., 2002):

- In contrast to the NOAEL approach, the BMD method takes into account the shape of the dose-response relationship to a greater extent, and is not limited to being one of the experimental dose levels.
- The use of a lower confidence bound (BMDL) appropriately reflects the sample size of a study, i.e. larger studies tend to result in shorter confidence intervals and thus lower uncertainty.
- A BMD, and BMDL, can be calculated even on occasions when a NOAEL cannot be determined.
- The BMD, and BMDL, corresponds to an explicit response level which introduces consistency and suggests that the POD for human health risk assessment will be based on more information relative to the NOAEL.

This last statement has been presented a major advantage, but it also represents one of the challenges associated with the benchmark methodology, i.e. which level of response, or risk, should the BMD correspond to?

Even though some early attempts were made to generalize the BMD concept (Crump, 1984; Gaylor and Slikker, 1990), initial discussions mostly concerned its application to quantal data (incidence data) with focus towards developmental toxicity (Faustman *et al*., 1994; Allen *et al*., 1994a and 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 Gaylor, 1988). Using a number of illustrative examples, Crump (1984) argued that according to this definition the BMDL will not strongly depend upon the model selected since the procedure does not involve extrapolation far below the experimental range. Apparently, the current EPA guidelines for determination of the POD in carcinogen risk assessment is similar to the original definition of the BMD(L) (USEPA, 2005a).

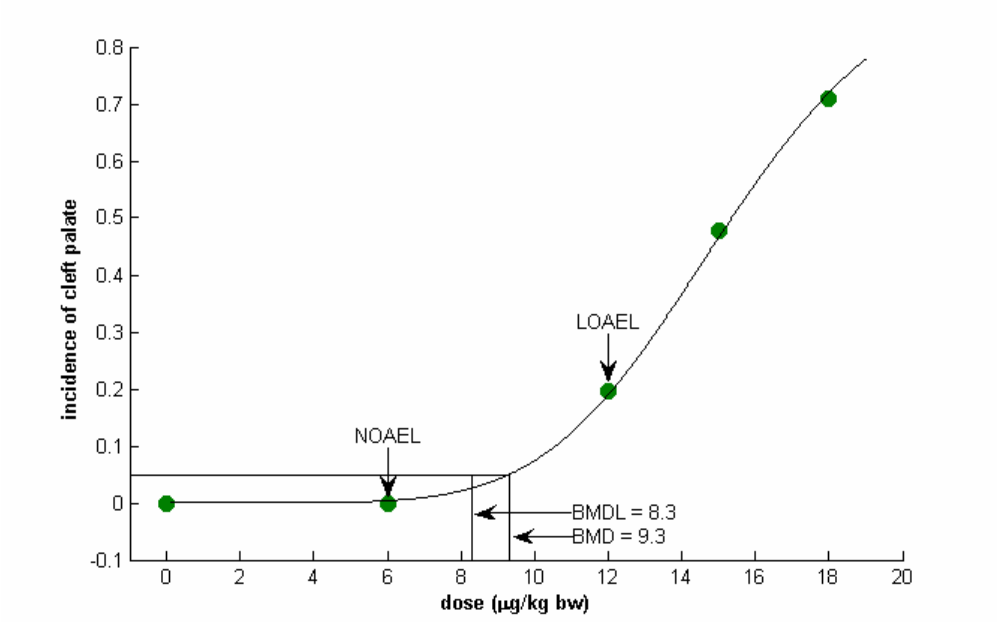

Figure 2. The BMD, and its lower bound (BMDL), defined as corresponding to an excess risk of 5%. The dose-response relationship is described by the log-logistic model. Incidence data on cleft palate observed mice fetuses following maternal exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) is used as basis (data from Birnbaum *et al*., 1989). The no-observed-adverse-effect level (NOAEL) and the lowest-observed-adverse-effect level (LOAEL) are 6 and 12 µg/kg bw, respectively.

While the BMD concept was readily introduced for quantal data its application to continuous dose-response information has warranted 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; **Papers II**, **III**, **IV**, and **V**). In contrast to quantal data where experimental subjects are categorized as responders or non-responders, for continuous data the severity of response is observed in the individual subject. Typical continuous responses constitute changes in organ weights, and enzyme activities. In the scope of originally defining the BMD as the dose corresponding to some level of risk for adverse health 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). In addition, alternatives to such formulations have been presented (Crump, 1984; Murrell *et al*., 1998; Slob and Pieters, 1998). Quantal data was initially considered in this thesis, but modeling of continuous dose-response data has been the major focus. In the following sections the different models and procedures suggested for BMD calculations will be discussed more in detail.

1.2.1 Quantal dose-response information

1.2.1.1 Dose-response models

For quantal data, a number of dose-response models have been suggested in health risk assessment of chemicals. These models are summarized 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 general, selection of the model to be used for BMD calculation has been suggested to be based on goodness-of-fit (ability to describe the data) (USEPA, 1995). In the present thesis, the models in Table 1 have been compared and procedures for model selection have been studied (**Paper I**).

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 intra-litter correlations occurs, meaning that the offspring from one litter respond more similarly than the offspring from another litter of a female in the same treatment group (Rai and Van Ryzin, 1985). This correlation problem has prompted 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 analyzed simultaneously have been presented (Krewski and Zhu, 1994 and 1995).

Table 1. Dose-response models for quantal data.

logistic model:

$$p(d) = \frac{1}{1 + e^{-(\alpha + \beta d)}}$$

log-logistic model:

$$p(d) = \gamma + \frac{1 - \gamma}{1 + e^{-(\alpha + \beta \ln d)}}$$

probit model[a]:

$$p(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha + \beta d} e^{\frac{-x^2}{2}} dx$$

log-probit model[a]:

$$p(d) = \gamma + \frac{1 - \gamma}{\sqrt{2\pi}} \int_{-\infty}^{\alpha + \beta \ln d} e^{\frac{-x^2}{2}} dx$$

Weibull model:

$$p(d) = \gamma + (1 - \gamma)\left(1 - e^{-\beta (d)^{\alpha}}\right)$$

gamma model[b]:

$$p(d) = \gamma + (1 - \gamma)\frac{1}{\Gamma(\alpha)} \int_{0}^{\beta d} x^{(\alpha - 1)} e^{-x} dx$$

multi-stage model:

$$p(d) = \gamma + (1 - \gamma)\left(1 - e^{-\sum_{j=1}^{n} \beta_j d^j}\right)$$

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

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

[b] In the model, $\Gamma(\alpha) = \int_{0}^{\infty} x^{(\alpha - 1)} e^{-x} dx$ is the gamma function.

1.2.1.2 Methods and applications of BMD analysis

As previously stated, the BMD, with lower bound (BMDL), was originally defined as the dose causing a 1 – 10% increase in the risk for adverse health effects compared to 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:

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

$$BMR = \frac{p(BMD) - p(0)}{1 - p(0)}, \text{ extra risk,} \quad (2)$$

where $p(BMD)$ is the probability of response at the BMD, and where $p(0)$ is the background probability of response. Apparently, the two definitions 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 and 1994b; Kavlock *et al*., 1995). In these studies the BMD method was evaluated and compared to 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 analyzed using the Weibull model. BMDLs (the lower 95% confidence bound of the BMD) corresponding to additional risks of 1, 5, and 10% were all found to be lower, on average, than the NOAELs, and the BMDLs associated with the 10% risk level most resembled the NOAELs. Models specifically designed for developmental toxicity testing were later applied. Using these models, BMDLs corresponding to an additional risk of 5% most resembled the NOAELs (Allen *et al*., 1994b).

The observations that a BMDL corresponding to an additional risk of 5 - 10%, on average, seems to be closest to the NOAEL for quantal data (Allen *et al*., 1994a and 1994b) have later been used as a basis for the BMR selection. This is exemplified in Allen *et al*., (1996) where BMDs were calculated for boric acid based on developmental toxicity observed in rats. Additionally, the same rationale was applied by Allen *et al*., (1998) in BMD calculations for developmental toxicity observed in rats after exposure to isopropanol. Recently, an inhalation reference concentration was proposed for methanol based on developmental data (Starr and Festa, 2003). In these calculations, the BMDL (the POD in their assessment) was defined as the 95% lower confidence limit of the dose causing an extra risk of 10%. The extra risk definition was used, rather than additional risk, due to a high background incidence, and a 10% level was used with the argument that a 5% change was below the observable response range.

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, 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 detailed comparison between BMDLs and NOAELs has been conducted for developmental toxicity, such analysis is fairly 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 database similar findings has later 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 between 1 – 10% were, on average, lower that the NOAELs (and $BMDL_{10}$ was, on average, closest to the NOAEL). However, for the acute inhalation data the differences between BMDLs and NOAELs were smaller.

An important aspect of the BMD approach is whether current experimental study designs, typically including a control and three treatment groups, are appropriate for BMD/BMDL estimation. Considering quantal data and developmental toxicity, the influence of study design has been analyzed by Kavlock *et al*., (1996), and focused on the accuracy and precision of a BMD corresponding to a likely target additional risk of 5%. The general conclusion was that the standard designs are adequate for benchmark calculations but may be improved by minor modifications. More recently, this issue was investigated by Krewski *et al*., (2002), also focusing on developmental toxicity and an excess risk of 5%. It was concluded that designs consisting of three dose groups, i.e. an untreated control and two treated groups, were close to optimal. However, suboptimal designs including more dose groups are often preferable since construction of the near-optimal design requires *a priori* knowledge of the underlying dose-response curve (Krewski *et al*., 2002).

1.2.2 Continuous dose-response information

1.2.2.1 Dose-response models

For continuous data, a large number of models may be used to describe the relationship between the dose of chemical and the mean response, $\mu(d)$. Dose-response models for continuous data that have been encountered in literature relating to the BMD method are presented in Table 2. The polynomial model and the power model have typically been used in previous applications of modeling continuous endpoints (Crump, 1984; Allen *et al*., 1994a, 1996, and 1998; Kavlock *et al*., 1995; Barton and Das, 1996; Haber *et al*., 1998; Gephart *et al*., 2001; Crump, 1995; Crump, 2002).

Table 2 also includes the Hill function, which is probably the most commonly used dose-response model in pharmacology, and it has also received attention in more recent applications relating to risk assessment of chemicals and BMD analysis (Barton *et al*., 1998; Murrell *et al*., 1998; Zhou *et al*., 2001 and 2002; Kim *et al*., 2002; Gaylor and Aylward, 2004; Toyoshiba *et al*., 2004; USEPA, 2005b; **Papers III**, **IV**, and **V**). A generalized form of this model known as the Richards function (Richards, 1959) is also given in Table 2. In addition, an exponential model was considered in this thesis (Table 2). This latter class of models has also been discussed and applied by Slob (2002), and a similar exponential model has been presented by Kalliomaa *et al*., (1998).

Table 2. Statistical models for continuous data.

Dose-response models	**Variance models**
polynomial model: $\mu(d) = \alpha + \beta_1 d^1 + \beta_2 d^2 + \ldots + \beta_n d^n$	
power model: $\mu(d) = \alpha + \beta d^\eta$	**constant model**: $\sigma^2(d) = c$
Hill function: $\mu(d) = \alpha + \theta \frac{d^\eta}{\kappa^\eta + d^\eta}$	**power function of the mean**: $\sigma^2(d) = \lambda[\mu(d)]^\rho$
Richards function: $\mu(d) = \alpha + \theta \left[\frac{d^\eta}{\kappa^\eta + d^\eta} \right]^\nu$	**dose dependent exponential model**: $\sigma^2(d) = e^{\lambda + \rho \ln(d+1)}$
exponential model: $\mu(d) = \alpha + \theta \left[1 - e^{-\left(\frac{d^\eta}{\kappa^\eta}\right)} \right]$	

Note: In the models, *d* represents the dose; and α is a parameter that describes the background response. The degree of the polynomial model is denoted by *n*. In the Hill function, the Richards function, and the exponential model, θ is a parameter that describes the magnitude, or size, of response; κ is the location parameter (which equals the ED_{50} in the Hill model); and η is a parameter that describes the shape of the dos-response curve. The Hill function is a special case of the Richards function, for $\nu = 1$, and is symmetrical on the log-dose scale. In the Richards function, for $\nu \neq 1$ the dose-response curve will have asymmetrical properties. Note that parameters κ and η do not have identical interpretation across the different models. This is also the case for parameters λ and ρ that define the non-constant variance models.

In contrast to the polynomial and power models, the Hill model, the Richards function, and the exponential model can produce S-shaped dose-response relationships, i.e. curves that level off to some maximum or minimum response level. The Hill function, which is illustrated in Figure 3, is the model most frequently considered in this thesis. Each of the parameters defining this function can easily be interpreted; α describes the background response; θ describes the magnitude, or size, of response (the difference between the maximum and minimum response level); κ is the location parameter, which equals the ED_{50}; and η describes the shape of the dose-response curve (η is usually called the Hill coefficient). A similar interpretation can be used for the parameters in the exponential model, except that the location parameter, κ, does not equal the ED_{50}.

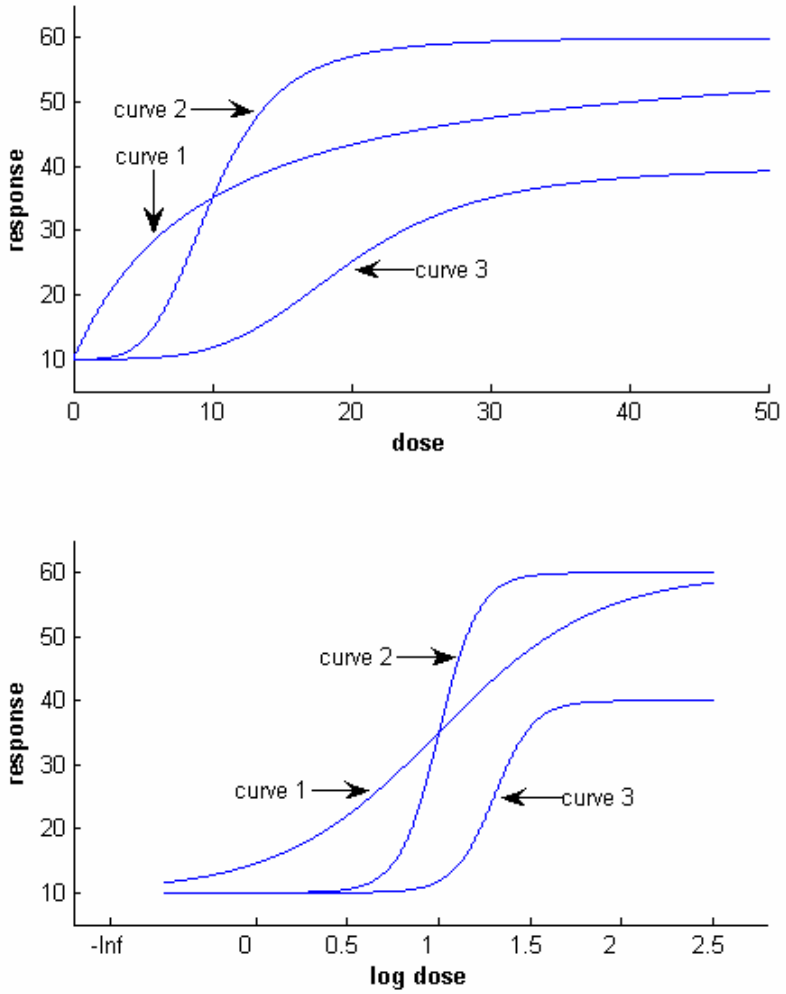

Figure 3. An illustration of the Hill function where the background, α, equals 10 for all cases. The magnitude of response, θ, equals 50 for curves 1 and 2, and it equals 30 for curve 3. The location parameter, κ, equals 10 for curves 1 and 2, and it equals 20 for curve 3. The Hill coefficient, η, equals 1 for curve 1, while it equals 4 for curves 2 and 3. The Hill function has symmetrical properties on the log-dose scale.

For a value of the Hill coefficient, η, equal to 1, the Hill function coincides with the Michaelis-Menten equation suggested to describe enzyme action (Cornish-Bowden, 1995). As shown in Figure 3, for $\eta = 1$ the Hill model cannot produce S-shaped curves on normal dose scale; it has initially the highest slope and successively levels off to some limiting response value as the dose increases. However, considering the log-dose scale the Hill model has always symmetrical properties (Figure 3). Asymmetrical characteristics of the dose-response relationship may be analyzed using the Richards function. In this model, for values of the ν parameter higher or lower than 1 the dose-response curve will have an asymmetrical structure (on the log-dose scale).

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 vise versa. Due to such observations, a model for the variance may be specified as a power function of the mean response (USEPA, 2005b). Of course other alternatives can be used for modeling the variance. In the present thesis, a dose dependent exponential model was also considered (Table 2).

1.2.2.2 Methods and applications of BMD analysis

A number of different procedures have been presented for how to calculate the BMD for continuous endpoints. The present thesis aims as an important part to analyze these procedures, which are herein are divided in two different categories depending on the resulting interpretation of the BMD:

Non-probability based interpretation of the BMD: Procedures where the BMD is defined as corresponding to a change in response relative to the mean (i.e. defined as a function of the mean response model), or the case where the BMD is defined as corresponding to a change in response relative to the standard deviation in the control group.

Probability based interpretation of the BMD: Procedures where the BMD is defined as corresponding to a change in the probability of "adverse" response, in resemblance to the case for quantal data.

1.2.2.2.1 Non-probability based interpretation of the BMD

The BMD for continuous data was in the early stages 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:

$$cBMR = \frac{|\mu(0) - \mu(BMD)|}{\mu(0)}, \qquad (3)$$

where $\mu(0)$ is the background mean response, and where $\mu(BMD)$ is the response at the BMD. In the present thesis, 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 given endpoint that is considered acceptable on the level of the individual organism (Slob, 2002). Thus, the CES should ideally be endpoint-specific. However, currently it seems not possible to reach consensus regarding critical effect sizes for common toxicological effect parameters (Dekkers *et al*., 2001).

As another alternative, the BMD has been defined as corresponding to a change in response relative to the standard deviation in the control group, $\sigma(0)$ (Crump, 1984; Crump, 1995; Kavlock *et al*., 1995):

$$cBMR = \frac{|\mu(0) - \mu(BMD)|}{\sigma(0)}. \quad (4)$$

This definition may also be classified in the other category of procedures suggested for continuous data, i.e. resulting in a probability based interpretation of the BMD (Crump, 1995). 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 affected fetuses in each litter (Allen *et al*., 1994a). This response was analyzed using a cBMR defined according to equation (3). A BMDL corresponding to a cBMR of 5% was, on average, closest to the NOAEL. 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, BMDLs corresponding to a cBMR of 5%, according to definition (3), most resembled the NOAELs. Considering definition (4) the same was the case for a cBMR of 0.5.

As was the case for quantal data, these observations have been used as a rationale for specifying values of the cBMR in other studies (Allen *et al*., 1996 and 1998). The BMD suggested for boric acid in Allen *et al*., (1996) was for example based on a 5% change in response relative to background. Gephart *et al*., (2001) specified the BMD as the dose causing a 10% change in the mean response. This was performed using the somewhat weak assumption that a cBMR of 10% would correspond to an additional risk of 10%, i.e. it was assumed that risk was directly proportional to the percentage change in the continuous variable. In the present thesis, the definition of the BMD as the dose causing a 5 and 10% change in response relative to background has also been used in an application to developmental neurotoxicity, i.e. data on spontaneous behavior using 2,2',4,4',5-pentabromodiphenyl ether (PBDE99) as a model substance (**Paper III**).

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

$$cBMR = \frac{|\mu(0) - \mu(BMD)|}{Magnitude}. \quad (5)$$

This way of expressing response is quite commonly adopted in dose-response analysis when models that include a parameter describing the magnitude of response are used (Table 2). Murrell *et al*., (1998) suggested the use of this definition since it may provide means of standardization for continuous data, and pointed out that this way of defining response is similar to the extra risk definition, i.e. equation (2). Murrell *et al*., (1998) argued that continuous responses may differ in terms of their magnitude of response,

which is accounted for in definition (5). Thus, a certain response change relative to this magnitude may be applicable across different continuous endpoints, in contrast to the case for the CES which is suggested to be endpoint specific. The expression of response according to equation (5) has for example been employed when calculating BMDs for TCDD based on biochemical responses (Kim *et al*., 2002), and it has also been considered in the present thesis (**Papers IV** and **V**).

The issue of study design in BMD calculations has to some extent been discussed for continuous endpoints. Considering the case of defining the BMD as corresponding to a CES of 5%, Slob *et al*., (2005) recently concluded that the performance of a design is first of all 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. Dose placement also turned out to be a crucial factor, and to minimize the risk for inadequate dose placement the use of multiple dose studies is favorable (Slob *et al*., 2005). Similar findings have also been observed at our department for the case of defining the BMD as corresponding to a 5% change in response relative to the magnitude of response (Kuljus *et al*., manuscript in preparation).

1.2.2.2.2 Probability based interpretation of the BMD

The other category of procedures for defining the BMD for continuous endpoints focuses on making statements in terms of probability (or risk), in similarity to the case for quantal data, thereby generalizing 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 analyzed 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 approach was presented by Gaylor and Slikker (1990), which is illustrated in Figure 4. Importantly, this procedure involves estimating the distribution of the data. As is shown in Figure 4, 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, $p(d_i)$, equals

$$p(d_i) = \phi\left[\frac{c - \mu(d_i)}{\sigma}\right], \qquad (6)$$

for a response that decreases with increasing dose of chemical, where ϕ is the cumulative standard normal distribution function; c is the cut-off value; $\mu(d_i)$ is the mean response at dose d_i; and where σ 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 compared to background, according to the additional or extra risk definition.

The procedure of calculating the BMD outlined above is sometimes referred to as the "hybrid approach" (Crump, 2002); a terminology that will be adopted in the present thesis. The fact that this approach does not involve data transformation has been put forward as a major advantage, i.e. theoretical simulation studies have shown that a higher precision is obtained in the BMD using the hybrid approach compared to procedures involving dichotomization of continuous response variables (West and Kodell, 1999; Crump, 2002).

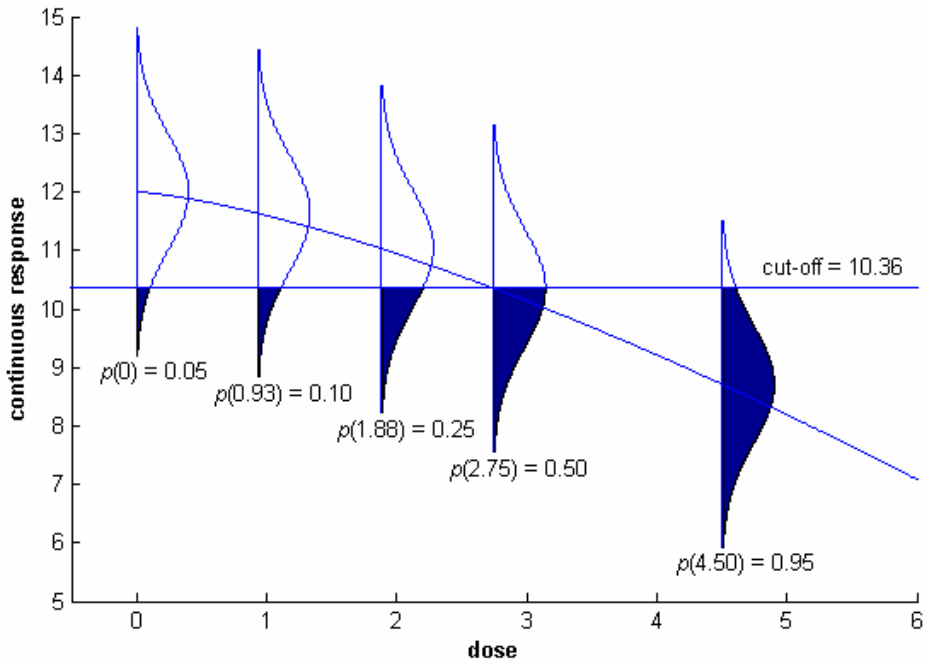

Figure 4. An illustration of the hybrid approach considering a hypothetical continuous response variable. The mean response as a function of dose is described by the power model (α = 12, β = -0.4, η = 1.4). The distribution of the data is assumed to be normal with standard deviation σ = 1.0. A cut-off 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 1 – 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).

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 estimated control distribution, i.e. corresponding to a certain percentile. Thus, the cut-off, *c*, in equation (6) may be solved as corresponding to some specified value of $p(0)$. Apparently, this procedure of specifying the cut-off fixes the background probability of "adverse response" as equal to the cut-off formulation. 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 and 2000).

Methodological aspects of the hybrid approach have been investigated (Kodell and West, 1993; West and Kodell, 1993 and 1999; Crump, 1995 and 2002). 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), illustrating the points of agreement of the two methodologies. In their draft benchmark dose technical guidance document (BMDTG), the USEPA discusses that the BMD can be defined as corresponding to a change in the mean equal to 1 control standard deviation, in the absence of any idea of what level is adverse. This indirectly suggests use of the hybrid approach with $p(0) = 0.02$, and BMR = 0.10 (USEPA, 2000). Methodological aspects of the hybrid approach have also been investigated in the present thesis (**Papers II** and **III**).

The hybrid methodology was initially illustrated for neurochemical and neurohistological effect data observed in monkeys and rats exposed to methylenedioxymethamphetamine (Gaylor and Slikker, 1990 and 1994). Further application of the approach to experimental neurotoxicity data has also been presented (Kodell *et al*., 1995; Slikker *et al*., 1998). In addition, in this thesis the hybrid model was applied to neurobehavioral endpoints (**Paper III**). Besides applications to neurotoxicity, 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 recently been compared to a case where the BMD was defined as corresponding to a 1% change in response relative to the magnitude of response, called the ED_{01} approach by the authors (Gaylor and Aylward, 2004). Gaylor and Aylward (2004) concluded that the hybrid procedure performed better than the ED_{01} approach, i.e. the uncertainty associated with a BMD corresponding to an additional risk of 1, 5, or 10% (estimated using the hybrid procedure) was lower than that associated with the ED_{01}.

While the use of human epidemiological data is not discussed in the present thesis, the use of the hybrid concept in such applications has gained some attention. The methodology is applicable to studies where the response and exposure has 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, and 2000), and also in association with exposure to polychlorinated biphenyls (Jacobson *et al*., 2002), and manganese (Clewell *et al*., 2003). The USEPA 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 are currently investigated at our institute using the hybrid procedure (Suwazono *et al*., manuscript in preparation). For epidemiological data, the interpretation of the BMD under the hybrid approach, i.e. as corresponding to a certain risk level, could also be reflected in a resulting guidance value since less extrapolation is required compared to the case for experimental data.

2 PRESENT INVESTIGATION

2.1 AIM OF THE THESIS

The aim of the present thesis was to investigate aspects of dose-response modeling and procedures for BMD analysis in health risk assessment of non-genotoxic chemical substances. While quantal data has been considered, the major focus has been directed towards continuous data. These two types of data will be discussed separately. A general objective concerned the question of which response the BMD should be associated with. In addition, for continuous data attention has been directed towards how the BMD should be defined. The specific objectives are given below:

Quantal data

Methodological aspects:

- To investigate the impact of model choice in BMD calculations, given the original definition of the BMD as the dose causing a 1 – 10% increase in the risk for adverse health effects compared to background, and to analyze procedures for model selection (**Paper I**).

Continuous data

Methodological aspects:

- To investigate different procedures for BMD analysis, including the hybrid approach (**Papers II** and **III**), the definition of the BMD as the dose causing a percentage change in response relative to background (**Papers III** and **IV**), and the definition of the BMD as the dose causing a percentage change in response relative to the magnitude, or size, of response (**Papers IV** and **V**).
- To discuss and propose developments in BMD analysis (**Paper V**).
- To address issues related to comparison of dose-response curves, and BMDs (**Paper IV**).

Applications:

- To suggest how the BMD concept can be applied for neurobehavioral endpoints, i.e. spontaneous behavior variables, which are observed over time (**Paper III**).
- To estimate strain differences in response to long-term TCDD exposure (**Paper IV**).

2.2 MATERIALS AND METHODS

An overview of the materials and methods used in the present thesis is given below. More detailed descriptions can be found in the individual **Papers I - V**.

2.2.1 Data material

Dose-response data was derived from the literature (**Paper I**) and obtained via scientific collaborations (**Papers III**, **IV**, and **V**). In **Paper I**, malformation data (incidence of cleft palate and hydronephrosis) observed in fetuses following exposure of pregnant C57Bl/6 mice to a number of different polychlorinated and polybrominated compounds, were collected from the scientific literature (Weber *et al*., 1985; Birnbaum *et al*., 1987a, 1987b, 1989, and 1991; Mayura *et al*., 1993). Since **Paper II** concerned theoretical investigations of the hybrid concept, no toxicity data was used. In **Paper III**, spontaneous behavior data (locomotion, rearing, and total activity) observed in 2-, 5-, and 8-month-old male and female C57Bl mice neonatally exposed to PBDE99 were obtained via collaboration with the Department of Environmental Toxicology at Uppsala University, Sweden. In contrast to other experimental data considered, this study was partly designed for purposes of dose-response modeling, i.e. it was a comparatively large study including a control and 5 treatment groups each consisting of 10 male and 10 female mice, respectively. In **Papers IV** and **V**, data on body and organ weights, altered hepatic foci, hepatic CYP1A1 induction, as well as retinoid and bone parameters, observed in female Long-Evans (L-E) and Han/Wistar (H/W) rats following long-term exposure to TCDD, were made available via collaboration within our department and with the Department of Environmental Health at the National Public Health Institute, Kuopio, Finland. Parts of this data base have previously been published (Viluksela *et al*., 2000; Jämsä *et al*., 2001; Fletcher *et al*., 2005).

2.2.2 Statistical methods

2.2.2.1 Model fitting

The maximum likelihood approach was used in the present thesis. This method requires that an assumption of the distribution of the data is made. For quantal endpoints, the probability of response in each dose group was assumed to be binomial. The log-likelihood function, ln L, then takes the form:

$$\ln L = \sum_{i=1}^{k} \left[\ln \binom{n_i}{x_i} + x_i \ln p_i + (n_i - x_i) \ln(1 - p_i) \right], \qquad (7)$$

where k is the number of dose groups; n_i is the total number of animals in the *i'th* dose group; and x_i is the number of responding animals in the *i'th* dose group. The parameters that define p_i are estimated by the values that maximize ln L.

Continuous endpoints were assumed to be normally distributed. For continuous dose-response data, which are normally distributed with mean, $\mu(d_i)$, and non-constant variance, $\sigma^2(d_i)$, over the dose levels, d_i, the log-likelihood function, ln L, takes the form:

$$\ln L = -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{g}\left[\frac{n_i}{2}\ln\sigma^2(d_i) + \frac{(n_i-1)s_i^2}{2\sigma^2(d_i)} + \frac{n_i(\bar{y}_i - \mu(d_i))^2}{2\sigma^2(d_i)}\right], \quad (8)$$

where N is the total number of animals; g is the number of dose groups; n_i is the number of animals in the *i'th* dose group; s_i^2 is the unbiased sample variance in the *i'th* dose group; and $\bar{y}_i$ is the mean response observed in the *i'th* dose group. The parameters defining the mean response, $\mu(d_i)$, and the variance, $\sigma^2(d_i)$, are estimated by the maximization of ln L.

2.2.2.2 Dose-response models

In **Paper I**, all the models presented in Table 1 were employed. In **Paper II**, a linear model was used for theoretical calculations. In **Paper III**, modified versions of the Hill function and the exponential model in Table 2 were used. In **Paper IV**, the Hill function was applied, and in **Paper V**, the Richards function and its special cases were considered. In the case of continuous dose-response data, a model for the variance also needs to be specified. A constant variance assumption was applied in **Papers II** and **III**. In **Papers IV** and **V** two non-constant models were considered in addition to the constant assumption; a power function of the mean response, and a dose-dependent exponential model (Table 2).

2.2.2.3 Analysis of model performance

For quantal data, model fit was analyzed using the Pearson χ^2 test statistics. The appropriateness of using the Akaike's information criterion (AIC) for model selection was also investigated. For continuous data, a likelihood ratio test statistic was used to assess individual model fit as well as to compare different models (which belong to the same class). 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 χ^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 a default in this type of testing, i.e. if $p \leq 0.05$ the two models were considered to be significantly different.

2.2.2.4 Calculation of confidence intervals

The likelihood ratio test statistics was, in addition to purposes associated with model comparison, used to calculate an approximate lower 95% confidence limit of the BMD, i.e. the BMDL, in **Papers I**, **III**, and **V**. The BMDL was defined as the dose satisfying

$$-2\left(\ln L_{rep} - \ln L\right) = \chi^2_{1,0.95},$$

where ln L is the value of equation (7) or (8) at the maximum likelihood estimates of the model parameters, and where ln L_{rep} is the value of equation (7) or (8) at the maximum likelihood estimates of the model parameters given a constraint on the BMD. In **Paper IV**, a likelihood ratio test statistic was similarly used for constructing a 95% confidence interval for a BMD ratio, describing intra-species sensitivity differences.

2.2.2.5 Simultaneous analysis of data

For continuous endpoints analyzed in this thesis, data were available for different sub-populations, i.e. for male and female mice in **Paper III**, and for L-E and H/W rats in **Paper IV**. In **Papers III** and **IV** models were fitted simultaneously to data from both sexes and rat strains, respectively. For this case the log-likelihood function, i.e. equation (8), needs to be slightly modified so that

$$\ln L = -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{2}\sum_{j=1}^{g_i}\left[\frac{n_{ij}}{2}\ln\sigma^2\left(d_{ij}\right) + \frac{\left(n_{ij}-1\right)s_{ij}^2}{2\sigma^2\left(d_{ij}\right)} + \frac{n_{ij}\left(\bar{y}_{ij} - \mu\left(d_{ij}\right)\right)^2}{2\sigma^2\left(d_{ij}\right)}\right],$$

where index i denotes the sub-population. Simultaneous analysis of data allows formal assessments (using likelihood ratio tests) of which model parameter/s (if any) may depend on sex, or strain. Thus, this type of analysis can be more effective since it may motivate a reduction in the number of parameters used for describing the data.

2.2.2.6 Software

In **Paper I**, the USEPA benchmark dose software (BMDS) was used (USEPA, 2005b). This software can be downloaded at http://cfpub.epa.gov/ncea. All the quantal dose-response models in Table 1 are available in the BMDS. In **Papers II**, **III**, **IV**, and **V**, the mathematical and statistical procedures associated with dose-response modeling and BMD calculations were established in Matlab (version 7.0).

2.3 RESULTS AND DISCUSSION

2.3.1 Quantal data

According to the initial presentation of the BMD method it was suggested that a BMDL corresponding to an additional or extra risk of 1 – 10% should not be very sensitive to the choice of model, since such a definition generally would not involve extrapolation far below the experimental range (Crump, 1984). In **Paper I**, this issue was investigated using the quantal dose-response models in Table 1. Malformation data, i.e. data on cleft palate and hydronephrosis, observed in mice fetuses following maternal exposure to polychlorinated and polybrominated compounds were used as the basis for the analysis. Methods and models specifically designed for developmental toxicity testing were not considered due to a lack of litter specific information in the studies from which the data were derived.

For the data on cleft palate, it was concluded that the BMDL was not sensitive to the choice of model at extra risks of 5 and 10% (Figure 5). The degree of model dependence was defined as the ratio between the highest and the lowest BMDL, i.e. $BMDL_{highest}/BMDL_{lowest}$. In the analysis, only models that satisfied global goodness-of-fit criteria, $p > 0.05$, as well as residual specific goodness-of-fit requirements were considered. At extra risks of 5 and 10%, $BMDL_{highest}/BMDL_{lowest}$ was below a factor 2 for all the data on cleft palate. In the draft BMDTG, the USEPA suggests that model dependence may not be considered apparent if $BMDL_{highest}/BMDL_{lowest}$ is within a 3-fold factor (USEPA, 2000). At the 5 and 10% levels of extra risk, the degree of model dependence was also similar between the different data sets suggesting that $BMDL_{highest}/BMDL_{lowest}$ was not very sensitive to the type of study design used (Figure 5). For example, the number of dose groups employed differed between the data sets. However, at an extra risk of 1% model dependence was more pronounced, and also more dependent on the particular data set considered. The 5% risk level was suggested for cleft palate since the choice of model, for BMDL calculation, did not appear to be important down to this risk level (Figure 5).

In **Paper I**, procedures for model selection were also discussed. The draft BMDTG suggests use of the AIC for selection between models which are not discharged under global goodness-of-fit criteria ($p > 0.1$), and that produce BMDLs that differ within a 3-fold range (if BMDLs between the models differ more, the most conservative BMDL may be selected instead) (USEPA, 2000). The use of AIC was investigated as a tool for model selection in cases when models could not be discriminated under global and residual specific goodness-of-fit criteria. It was concluded that use of the AIC for selection between dose-response models available in the BMDS was problematic. The AIC values were on several occasions rather similar between the models for a given data set. In such cases, model ranking according to the AIC may not be very significant (Burnham and Anderson, 1998).

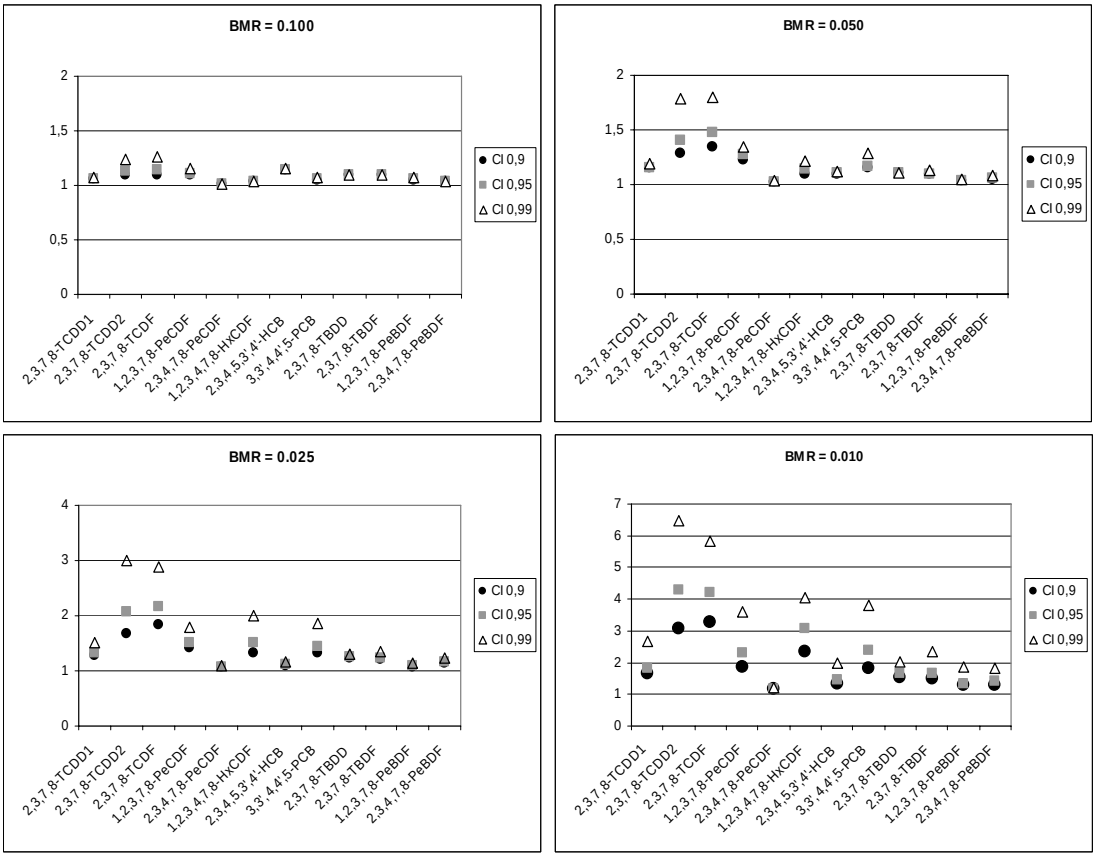

Figure 5. Ratios between the highest and the lowest BMDL ($BMDL_{highest}/BMDL_{lowest}$) obtained using the quantal models in Table 1. The ratio, $BMDL_{highest}/BMDL_{lowest}$, is presented for different risk levels (BMRs) and confidence intervals (CIs) defining the BMDL. Incidence data on cleft palate is used as a basis (Weber *et al*., 1985; Birnbaum *et al*., 1987a, 1987b, 1989, and 1991; Mayura *et al*., 1993). Only models that satisfied global goodness-of-fit criteria, $p > 0.05$, as well as residual specific goodness-of-fit requirements were considered for a given data set.

While the draft BMDTG suggests the use of AIC in cases when several models adequately fit the data, it states that procedures of model selection at this stage may be regarded as somewhat arbitrary (USEPA, 2000). Thus, a conservative model selection approach was considered more appropriate in **Paper I** (the model with the lowest BMDL was selected), which consistently resulted in the use of the multi-stage model for the data on cleft palate at extra risks of 5% and below. The quantal dose-response models (Table 1) may be quite easily graded according to their conservatism at lower response levels, when model dependence becomes more pronounced. The log-probit model may be regarded as the least conservative model; it has the most "threshold-like" characteristics in the low-dose region while the opposite is the case for the multi-stage model (Figure 6). This was also observed by Krewski and Van Ryzin (1981). A conservative model selection approach is generally suggested in the present thesis (in cases when several models adequately describe the data) unless a clear preference for a certain model can be argued.

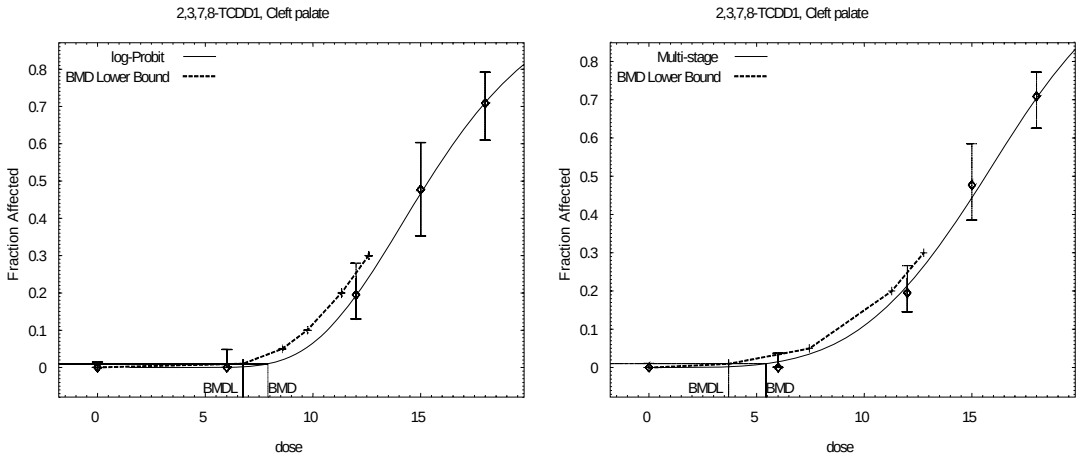

Figure 6. The log-probit model (left) and the multi-stage model (right) fitted to incidence data on cleft palate observed mice fetuses following maternal exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) (data from Birnbaum *et al*., 1989). The BMD, and BMDL, corresponds to an extra risk of 1%.

2.3.2 Continuous data

2.3.2.1 The hybrid approach

In **Paper II**, the hybrid approach was investigated from a theoretical point of view. Previous studies have compared the hybrid approach with procedures for dichotomization of continuous data (West and Kodell, 1999; Crump, 2002). However, in this thesis the influence of variance on the hybrid model was analyzed.

It was concluded that the influence of variance on the hybrid model depends on how the cut-off value is defined. As is shown in Figure 7, 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. However, if the cut-off is defined more independently of the model, i.e. directly as some level of the continuous response variable, the BMD becomes biased upward when the variance is biased downward. The former way of determining the cut-off (i.e. in terms of a percentile) is what has been suggested in the literature, and represents a common way of defining the "normal" range in a population (Crump, 2002). While it may not be as straight forward to directly specify a cut-off, an option may be to estimate the cut-off in a statistical sense (i.e. in terms of a percentile) using a large historical control material, and then apply this cut-off, in terms of its continuous response value, in BMD calculations.

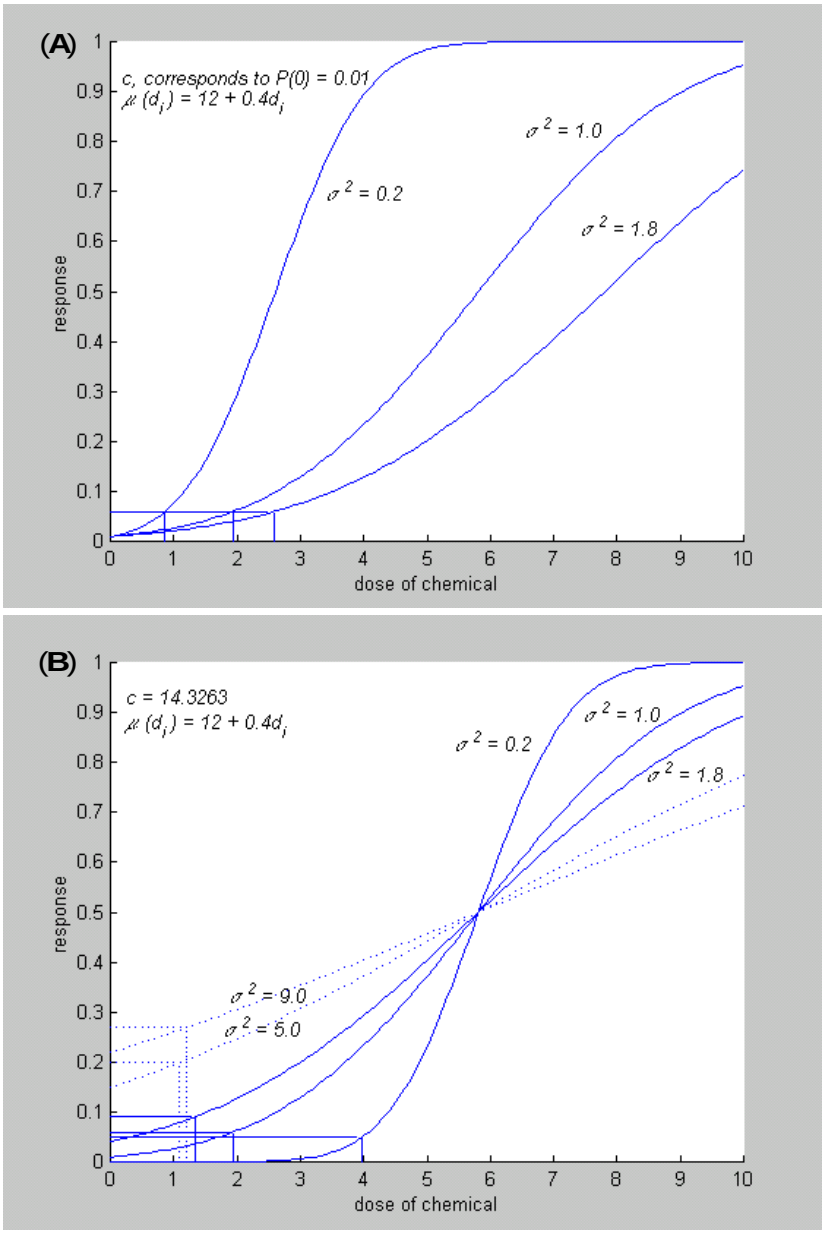

Figure 7. Probability dose-response models derived using the hybrid approach given a hypothetical continuous response variable which is normally distributed, and for which the mean response as a function of dose is described by a linear model, $\mu(d_i) = 12 + 0.4d_i$. Hybrid models resulting for different levels of variance, σ^2 (which is assumed to be constant over dose), are illustrated. In part A, the cut-off value, c, corresponds to the 99th percentile of the control distribution, $p(0) = 0.01$. In part B, c is fixed to a value of the continuous response variable ≈ 14.3263, which corresponds to $p(0) = 0.01$ for $\sigma^2 = 1.0$. BMDs corresponding to an additional risk of 5% are indicated.

In the context of the hybrid approach, problems of how to interpret the variance in experimental studies have been addressed (Murrell *et al.*, 1998; Slob, 2002; Gaylor and Slikker, 2004). Gaylor and Slikker (2004) recently discussed the importance of separating the inherent component of variance among animals and the variance of the experimental measurements. They concluded that if the overall standard deviation is

used, which may be biased upward, the risk at the BMD becomes underestimated. This can be explored in Figure 7A. However, considering the case of a model independent cut-off in Figure 7B, the opposite characteristic is generally obtained and such a property may be more appropriate in a risk assessment context. It should however be pointed out that simulation studies in **Paper II** suggested that, for low levels of additional risks (BMRs of 1, and 5%), a higher precision in the BMD was obtained when the cut-off was defined in terms of a percentile compared to the case of a fixed cut-off.

The hybrid approach introduces the concept of risk for continuous data without the need to perform dichotomization, and represents a standardized procedure of BMD analysis. However, the choice and definition of the cut-off value, and the selection of the BMR, are subject to discussion. The assumption of constant variance is used in the hybrid approach, but the case of non-constant variance is generally observed for experimental toxicity data. The applicability of the hybrid concept in the non-constant variance case has only been discussed to minor extent (West and Kodell, 1993). It has been pointed out that the hybrid approach may be complicated in cases when the variance in treated groups is lower than that in the control group, since this could result in that the probability of "adverse response" in treated groups becomes lower than in the control, i.e. an occurrence known as hormesis (Stiteler and Swartout, 1991).

2.3.2.2 Application of the BMD method to neurobehavioral endpoints

In **Paper III**, the BMD approach was introduced for neurobehavioral endpoints, i.e. data on spontaneous behavior observed in mice neonatally exposed to PBDE99. Following the theoretical studies in **Paper II**, the hybrid approach was applied to selected parts of the data. The definition of the BMD as the dose causing a percentage change in response relative to background was used as a general approach, applied to all the behavior variables (i.e. locomotion, rearing, and total activity), while the hybrid concept was also used for the data on locomotion.

An initial consideration of **Paper III** was how to quantify spontaneous behavior, which is observed over time. To resolve this issue, a response definition termed the "fractional response" was established. According to the fractional response, spontaneous behavior observed in the individual animal was defined as the cumulative motor activity after 20 minutes divided by cumulative activity during the whole 60-min test period (Figure 8). The fractional response contains information about the time-response profile, which differs between the treatment groups, and was considered to have appropriate statistical characteristics, i.e. it enabled the use of a common variance assumption according to a likelihood ratio test statistic. As previously stated, the case of constant variance is generally considered in applications of the hybrid method. While the fractional response definition appropriately accounts for the time factor, methods involving dose-time-response modeling for BMD calculations from neurobehavioral screening data (i.e. not spontaneous behavior) have recently been discussed (Zhu *et al*., 2005). However, if such methods are also suitable for modeling of spontaneous behavior is beyond the scope of the present thesis.

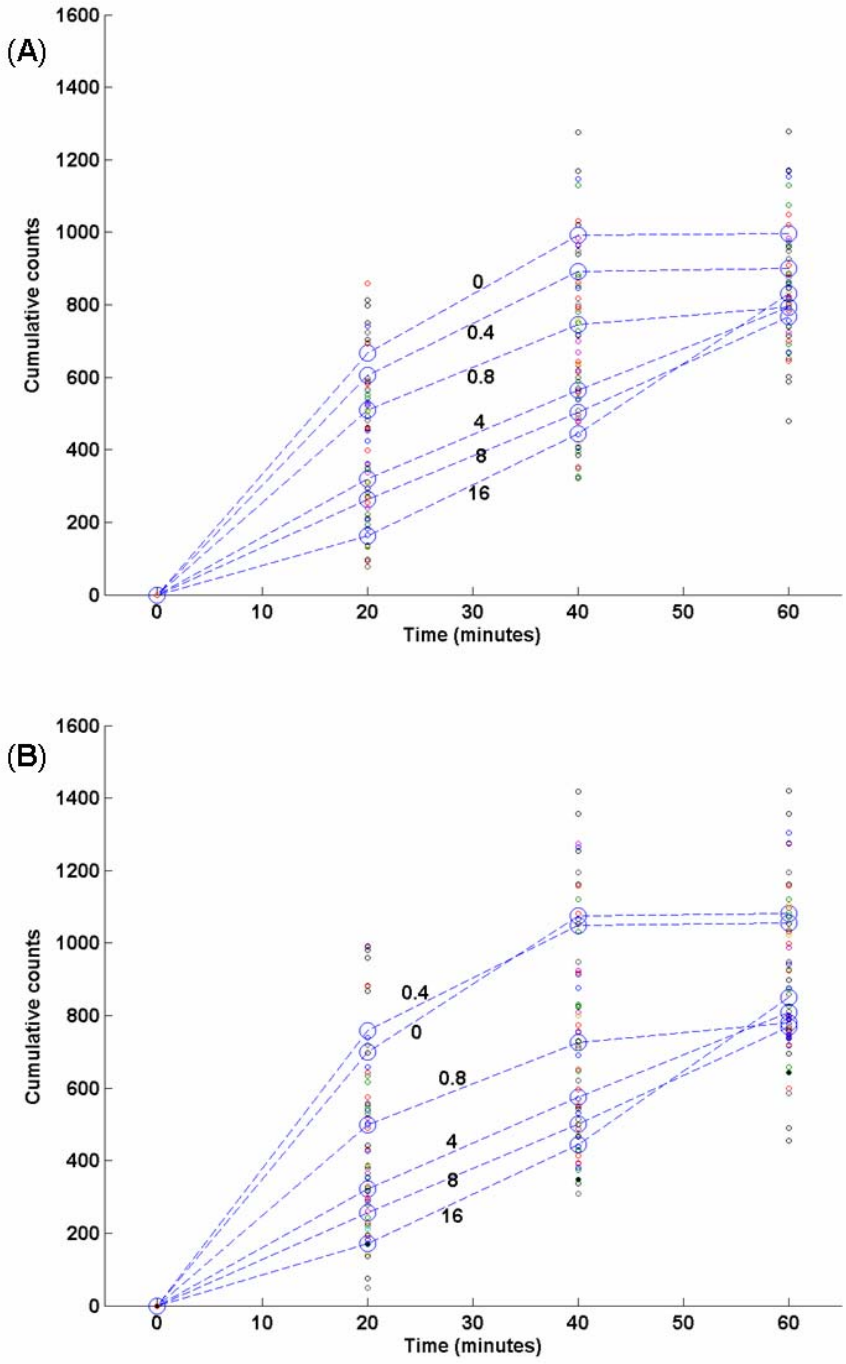

Figure 8. The cumulative number of counts for locomotion observed in 2-month-old male (part A) and female (part B) C57Bl mice orally exposed on postnatal day 10 to 0, 0.4, 0.8, 4, 8, or 16 mg PBDE99/kg bw. At each dose level, means (large circles) and individual observations (small circles) after 20, 40, and 60 minutes of testing (i.e. R_{20}, R_{40}, and R_{60}) are presented. Note that the dashed lines are used for keeping track of means resulting from the same treatment and are not estimates of time-response relationships. Individual behavior was quantified in terms of a fractional response; the number of counts produced after 20 minutes divided by the total number of counts produced over 60 minutes of testing, i.e. $100 \times R_{20}/R_{60}$.

Two different dose-response models, i.e. a modification of the Hill function and an exponential model, were used for modeling spontaneous behavior defined in terms of the fractional response. In Figure 9, both models fitted to data on locomotion are illustrated. According to statistical analysis, common dose-response model parameters could be used for male and female mice (considering each behavior variable) indicating that they responded similarly to the exposure to PBDE99. In addition, since the variance was not gender specific for locomotion and rearing, a rather simple model (including 4 parameters) could be established for these behavior variables using the complete body of data as a basis (i.e. 120 animals, 60 males and 60 females). Application of the hybrid concept was considered most appropriate for locomotion and rearing; the hybrid method probably benefits from using a large study group as basis. It has been pointed out previously that estimates of variance may be quite variable for experiments with small sample size while estimates of the continuous mean response can be more precise (Gaylor and Chen, 1996). Locomotion was selected for hybrid calculations since it is a quite common spontaneous behavior variable, and also because it may be regarded as a more specific measure of motor activity compared to rearing and total activity.

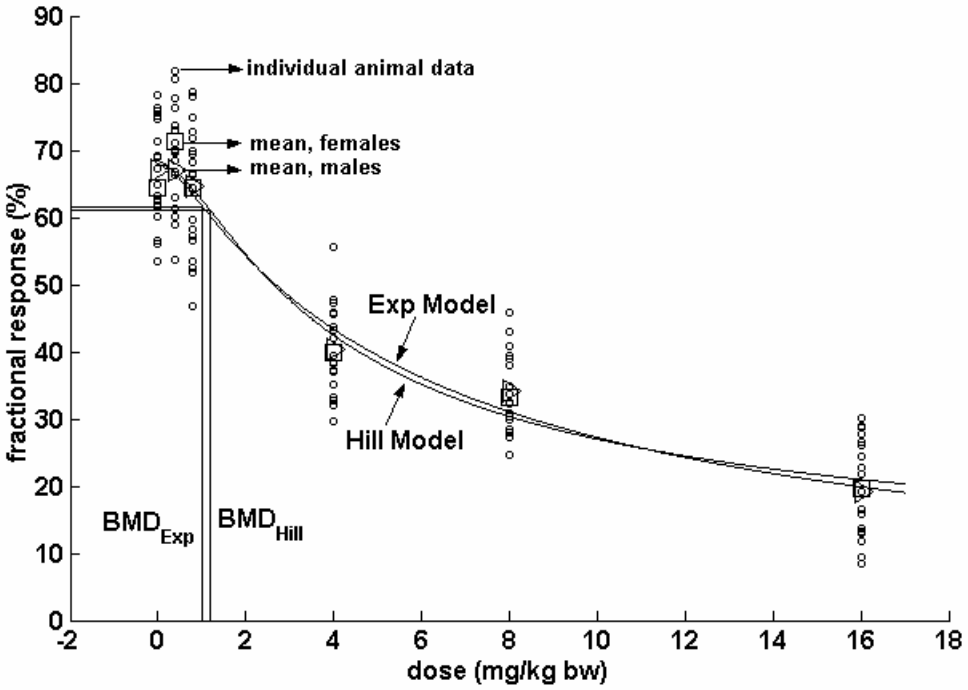

Figure 9. The Hill and Exp models fitted simultaneously to male and female data on locomotion observed in 2-month-old C57Bl mice neonatally exposed to PBDE99. BMDs corresponding to a 10% change in fractional response relative to background are shown. The mean of the fractional response for male (triangles) and female (squares) mice, as well as individual observations (small circles), are illustrated at each dose level.

In **Paper III**, the results indicated that BMDLs (and BMDs), corresponding to a 5 and 10% change in response relative to background, were similar between the two models, i.e. they generally differed within a 2-fold range. The NOAELs were most similar to BMDLs corresponding to a 10% change in response relative to background. Total activity was observed to be the most conservative (sensitive) endpoint. For a 5% change in response, the BMDL (according to the Hill model) for total activity was 0.15 mg/kg bw. For locomotion, the corresponding BMDL was 0.52 mg/kg bw. Considering the hybrid approach with a cut-off specified as the 5^{th} percentile of the control distribution, BMDLs for locomotion corresponding to additional risks of 5 and 10% were 0.45 and 0.63 mg/kg bw, respectively. Thus, for locomotion, the most common measure of spontaneous behavior, a BMDL of about 0.5 mg/kg bw was obtained for PBDE99 using "standard settings" within proposed methods of BMD analysis for continuous endpoints.

2.3.2.3 Comparison of dose-response curves: Strain differences in sensitivity

Assessment of strain and species differences in sensitivity for a given exposure, or calculation of relative potencies between chemicals, is ideally performed by comparing the corresponding dose-response curves. Since the BMD is derived from an estimated dose-response curve and is regarded to contain more information of the dose-response relationship compared to the NOAEL, the BMD also represents a quantity that better reflects the sensitivity of the species, or the potency of the chemical, relative to the NOAEL.

In **Paper IV**, the issue of comparing dose-response curves and BMDs was investigated considering dioxin sensitive L-E rats and dioxin resistant H/W rats following long-term exposure to TCDD. Sensitivity differences between L-E and H/W rats were investigated considering a number of toxicological endpoints including data on body and organ weights, altered hepatic foci, hepatic CYP1A1/A2 induction, as well as retinoid and bone parameters. While the two rat strains and their response to dioxin exposure have been studied extensively, a quantitative and statistical analysis of their relative difference in sensitivity following long-term TCDD exposure has prior to this work not been conducted.

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. $BMD_{H/W}/BMD_{L-E}$, estimated by the Hill function. The definition of the BMD as the dose causing a % change in response relative to background, as well as the definition of the BMD as the dose causing a % change in response relative to the magnitude of response was considered. It is important to note that the BMD ratio may not be independent of the level of response selected, which occurs 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, $BMD_{H/W}/BMD_{L-E}$ is constant and reflects the difference in the location of the curves on the dose scale. Thus, given the case of common dose-response patterns a sensitivity difference can be summarized in terms of a single quantity. The assumption of similar dose-response curves for L-E and H/W rats was generally supported

according to statistical analysis. A confidence interval for $BMD_{H/W}/BMD_{L-E}$ was also calculated denoting the uncertainty in the estimated sensitivity difference. Moreover, based on a likelihood ratio test statistic it was investigated whether, or not, the two strains differed statistically in their response to TCDD treatment; it was tested if $BMD_{H/W}/BMD_{L-E} = 1$, in which case the L-E and H/W dose-response curves have identical location.

In **Paper IV,** results demonstrated that L-E and H/W rats differed statistically in their response to TCDD treatment for parameters related to body and organ weights, altered hepatic foci, retinoid parameters, and hepatic CYP1A1 induction. The dose-response model estimated for body weight gain is illustrated in Figure 10. For bone effects significant strain differences were observed for tibia length and energy absorption but not for other bone parameters. Differences between the strains were most pronounced for volume fraction of altered hepatic foci, i.e. L-E rats were approximately 80 times more sensitive than H/W rats. For body and organ weight parameters, as well as tibia length and energy absorption L-E rats were approximately 10-20 times more sensitive than H/W rats. For the retinoid parameters and hepatic CYP1A1 induction, the quantitative estimates of strain differences in response to TCDD exposure were generally about 5-fold, and associated with a low uncertainty.

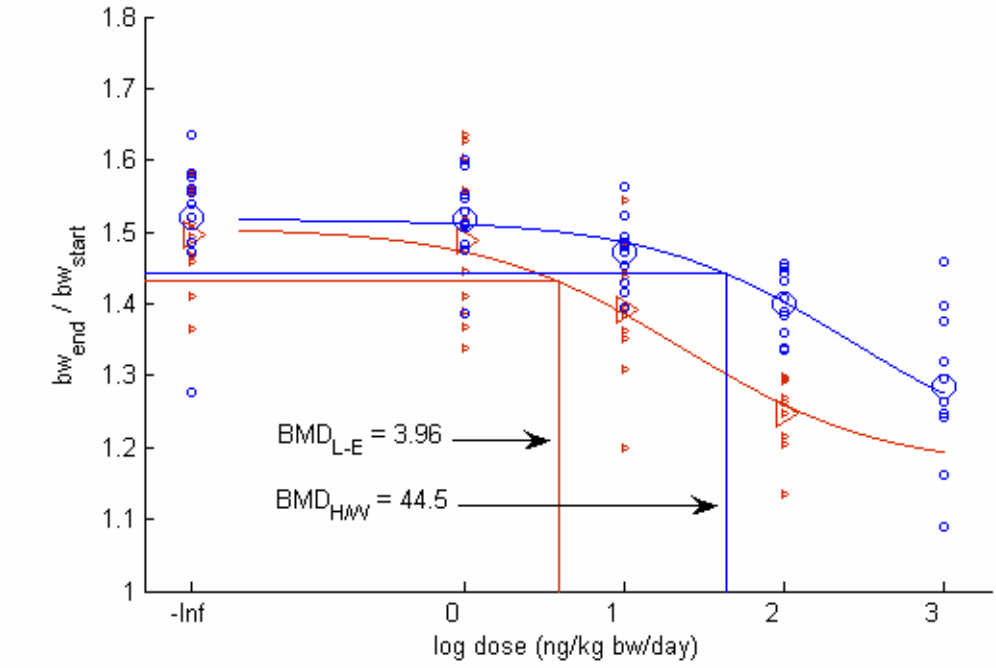

Figure 10. The Hill function estimated for data on body weight gain (bw_{end}/bw_{start}) observed in L-E (triangles) and H/W (circles) rats following long-term TCDD exposure. The BMD corresponds to a 5% change in response relative to background. The strain difference in sensitivity to TCDD exposure presented as the ratio, $BMD_{H/W}/BMD_{L-E}$ equals 11 (44.5/3.96). Since the dose-response curves are assumed to be similar for the two strains the BMD ratio is independent of the response level selected.

2.3.2.4 Developments in BMD analysis

In **Paper V**, developments in BMD analysis of continuous endpoints were discussed. This concerned the identification and estimation of a critical dose level in dose-response relationships that have S-shaped characteristics. The critical exposure of interest was defined as the dose where the slope of the curve changes the most in the low dose-region; a dose that may be considered to reside in a "transition dose range" where the sensitivity to chemical exposure may change noticeably. As is shown in Figure 11, the dose where the curve changes the most is the dose, denoted BMD_T, where the third derivative of the dose-response function equals zero. The consideration of dose-dependent transitions in mechanisms of toxicity has been argued as a way to improve the risk assessment process (Slikker *et al*., 2004a and 2004b). For example, the identification of a transition dose region is suggested to be a key to properly define the lowest range of doses appropriate for extrapolation between different species. Moreover, the region where a fundamental shift occurs in the dose-response relationship relates to the concept of a threshold dose. This has been discussed by Dybing *et al*., (2002) where it is argued that a biological threshold, in conformity to a transition dose range, represents a certain dose region where a substantial change in response occur, rather than a single dose level.

The identification of the BMD_T was performed considering a flexible mathematical model known as the Richards function (Richards, 1959). This function was considered since it very generally describes different possible characteristics that may be exhibited by an S-shaped dose-response curve. It can be shown that several of the common dose-response models in fact are special cases of the Richards function, including the Michaelis-Menten equation, the Hill function, the Gompertz curve, and the exponential model family (Table 2) (Richards, 1959; Giraldo *et al*., 2002). From a methodological point of view, the use of a BMD_T bypasses problems concerning which response the BMD should be associated with; rather both the degree of exposure and response resides within the proposed definition.

In **Paper V**, it was concluded that the response (if defined as a % change in response relative to the magnitude of response) associated with the BMD_T depends on the geometrical shape of the dose-response curve. More specifically, the response decreases as the curve becomes more asymmetrical with more threshold-like characteristics in the low dose region. Considering a symmetrical case, given by the Hill function, the response associated with BMD_T is approximately 21%. The Richards function has a limiting case, i.e. the Gompertz curve, as it becomes more and more threshold-like in the low dose region. According to the Gompertz curve, the response associated with BMD_T is 7.3%. In Figure 12, the BMD_T for both the Hill function and the Gompertz curve is illustrated. In cases when it is problematic to discriminate between the symmetrical and asymmetrical assumption, a conservative approach may be to use the limiting case response, i.e. the response associated with BMD_T according to the Gompertz curve (7.3%). The Gompertz curve, as well as other models, could then be considered for estimation of the corresponding dose. It was shown in **Paper V** that other models should generally perform conservatively relative to the Gompertz

curve since they are less threshold-like (for example the Hill function and the exponential model, Table 2).

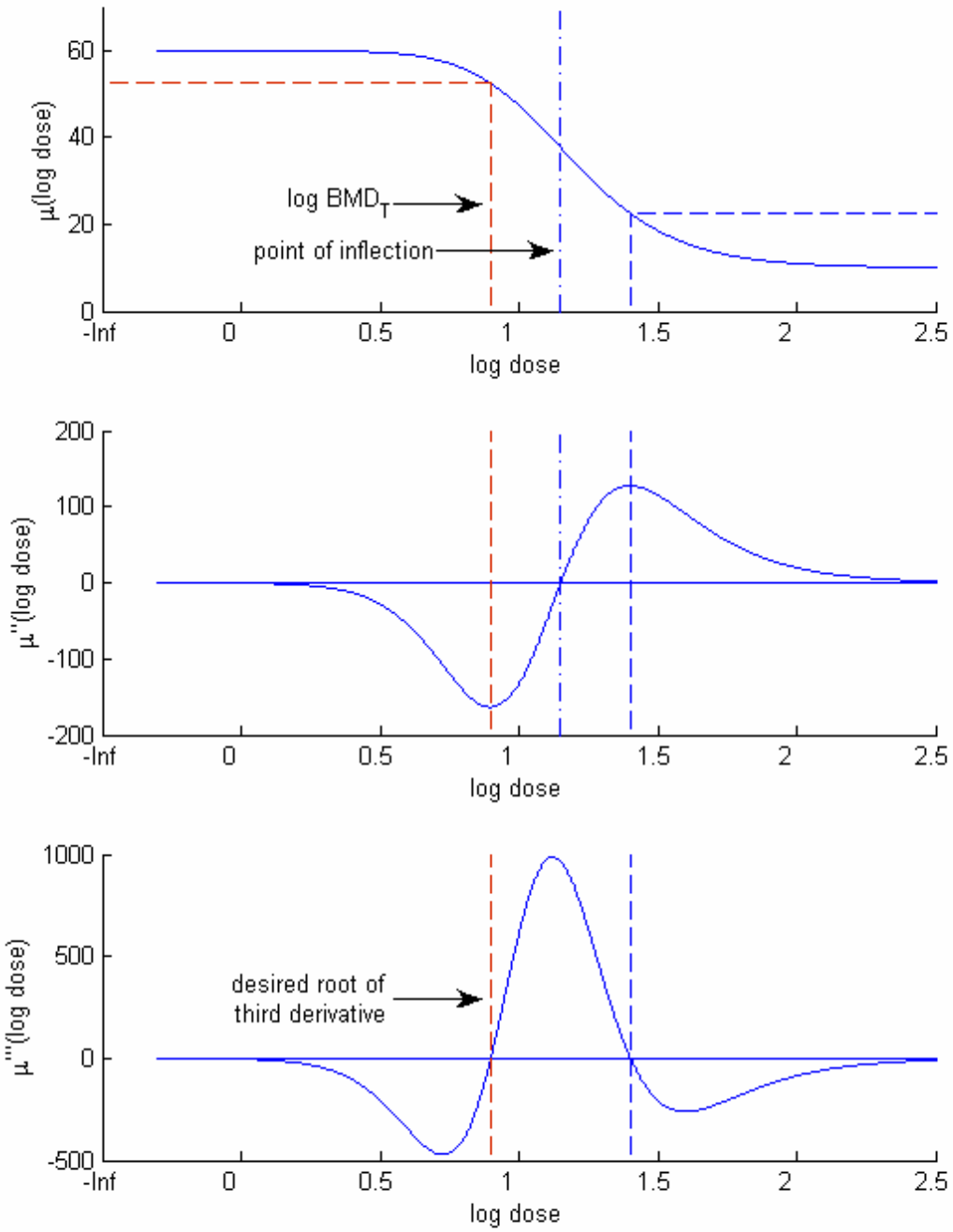

Figure 11. The Richards function and its second and third derivative. In the example, α = 60, θ = -50, κ = 10, η = 2, and ν = 2. The BMD_T is defined as the dose, in the low dose region, where the slope of the curve changes the most. As shown, this dose level corresponds to the dose where the second derivative is the lowest, and where the third derivative equals zero.

For estimating the BMD_T, or any other dose corresponding to a % change in response relative to the size of response, it is important to include high dose levels in the experiment so that information regarding the maximum or minimum response level is obtained. However, according to a recent study by Slob *et al*., (2005), this has also been shown to be of interest when defining the BMD in terms of a CES. If only a few treatment groups are used, in accordance with current OECD guidelines, it is probably important to have observations within the transition dose range, both in the low and high dose region, to accurately estimate the S-shaped dose-response curve. Figure 12 represents an example when this seems to be the case for an "OECD design" including a control and 3 treatment groups. As *a priori* knowledge concerning the shape of the dose-response curve generally is lacking an alternative is to distribute animals into several dose groups to avoid the risk for unfavorable dose placement (Slob *et al*., 2005; Kuljus *et al*., manuscript in preparation).

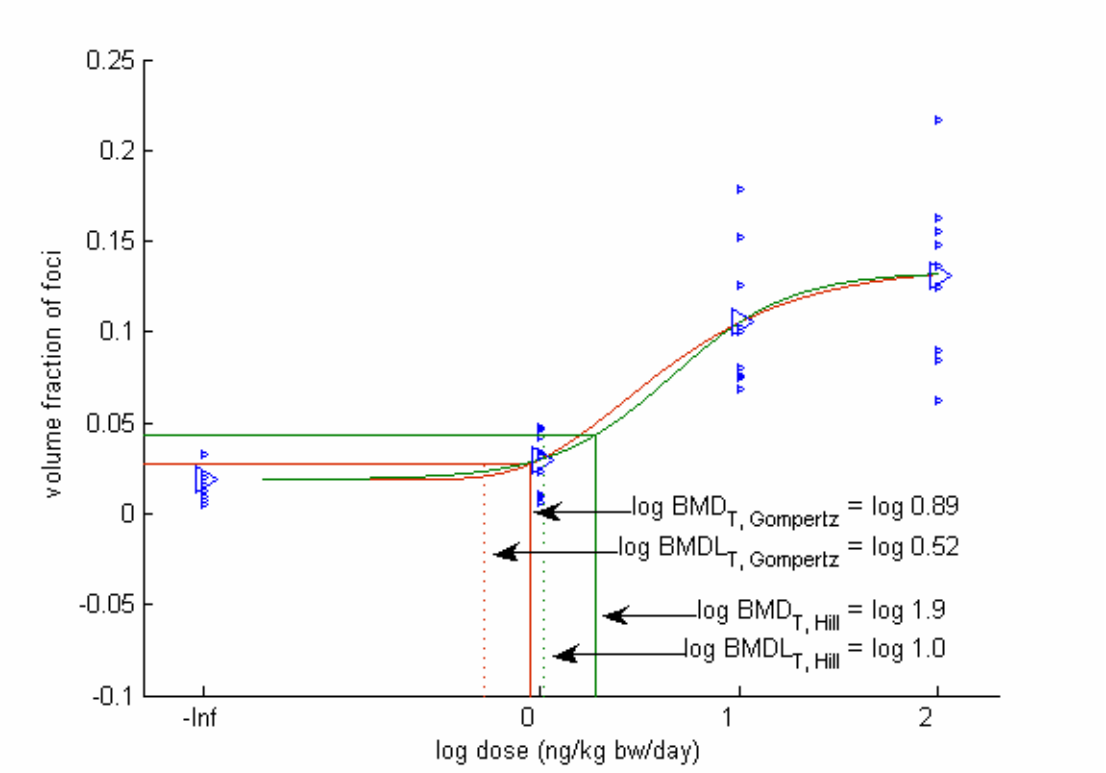

Figure 12. The Hill function and the Gompertz curve fitted to data on volume fraction of hepatic foci observed in L-E rats following long-term exposure to TCDD. The BMD_T (with lower bound $BMDL_T$) is defined as the dose where the slope changes the most. The response (defined as a % change in response relative to the magnitude, or size, of response) corresponding to the BMD_T is approximately 21% and 7.3% according to the Hill function and the Gompertz curve, respectively.

2.4 CONCLUSIONS AND FUTURE PERSPECTIVES

In this thesis, aspects of dose-response modeling and procedures for BMD analysis in health risk assessment of chemicals have been investigated. While the BMD method is used by authorities like the USEPA, this approach has not yet found its way into regulatory toxicology in Europe. However, while slowly progressing, discussion and awareness of the BMD concept has increased over the recent years within the EU. In the scope of this development, the present thesis provides an important input in the evaluation of the BMD approach that may help facilitating its introduction, thereby improving the health risk assessment of chemical substances.

A number of theoretical advantages of the BMD method have generally been used as arguments over the traditional NOAEL concept. Moreover, studies have indicated that the NOAEL should not be considered a risk-free exposure level, and the risk, or response, at the NOAEL may vary depending on the study. The fact that the BMD corresponds to an explicit response level has been argued as more appropriate since this introduces consistency and suggests that the point of departure for risk assessment will be based on more information. Indeed, this feature is a clear advantage but it also represents one of the challenges associated with the BMD concept, i.e. which response should the BMD correspond to? In addition, for continuous data the question of how the BMD should be defined is not straightforward. These issues have been addressed in the present investigation.

For continuous experimental data, the question of how to define the BMD itself needs further discussion. Problems are apparent with the definition of the BMD as the dose causing a percentage change in response relative to background, and also the hybrid approach. While the former case represents a simple way of defining the BMD, the development of endpoint specific response values may be complicated. The settings used within the hybrid concept, i.e. the cut-off value and the benchmark response level, are also subject to discussion. The application of the hybrid method in the present thesis emphasized some aspects of data requirements that are important, for example the case of common variance. On an overall basis, the definition of the BMD as corresponding to a critical effect size seems in general to be more applicable to experimental data compared to the hybrid approach.

While problems in BMD analysis of continuous endpoints may not have been resolved, developments were proposed. It is suggested that the BMD can be defined as the dose where a fundamental shift occurs in the S-shaped dose-response relationship. Such an exposure level can be specified as the dose where the slope of the curve changes the most in the low dose region. This approach may provide a biological basis for selecting the response associated with the BMD. Given a conservative scenario, it was shown that the response corresponding to the dose of interest is in the range between 5 - 10%, if defined as a percentage change relative to the magnitude of response. This definition of the BMD is suggested in future applications for continuous experimental data that contain information about the maximum or minimum response level.

For quantal data, in the absence of any preference for a particular response level, the use of a 5 or 10% risk level, rather than the 1% level, is suggested from a general point of view since this would reduce model dependence. Also, in comparison to the suggestions for continuous data, consideration of extra risks between 5 - 10% may provide a form of harmonization; while resulting BMDs for quantal and continuous data will not have identical meaning under the suggested definitions, the way of expressing the response change corresponding to the BMD, as well as the degree of response change, would be similar.

Some applications of dose-response modeling and BMD analysis have been presented in this thesis. As one part, the BMD method was introduced for neurobehavioral endpoints, i.e. spontaneous behavior, by the establishment of a response definition with appropriate statistical characteristics. As a second part, statistical differences in sensitivity between Long-Evans and Han/Wistar rats following long-term TCDD exposure were demonstrated for the first time. In this latter investigation, methodological issues regarding comparison of dose-response curves and BMDs were also addressed. The procedure discussed has broader applications in the assessment of species, strain, and sex sensitivity following chemical exposure.

In conclusion, this thesis discusses and illustrates several critical issues in dose-response modeling and BMD analysis. It also proposes a new BMD definition that might overcome some of the problems identified with the application of the BMD concept for continuous endpoints.

3 ACKNOWLEDGEMENTS

This investigation was performed at the Institute of Environmental Medicine (IMM), Karolinska Institutet, Stockholm, Sweden. Many people have been important during my doctoral studies, including colleagues as well as friends and family. To all of you, I am very grateful!

I would like to express my sincere gratitude to Dr. Agneta Falk Filipsson for being an excellent supervisor. Thanks for the great support and guidance that I have received from you during these years! And thanks for giving me responsibility and allowing me to explore my own scientific ideas.

Thanks to Associate Professor Katarina Victorin, my co-supervisor, for your many contributions and great encouragement! And I would also like to acknowledge your initiating role in this project.

Thanks to Professor Dietrich von Rosen, my other co-supervisor, for showing a lot of confidence and interest in this work! And thanks for always finding the time for fruitful discussion.

Thanks also to Professor Helen Håkansson, for your interest and support related to this project, and for establishing scientific collaboration and contacts in Europe.

Further, I would like to express my gratitude to:

David, for great friendship. Good luck with your thesis and other important "projects"!

Past and present members of the Division of Environmental Health Risk Assessment, who are a nice group of people to work an socialize with, including Nicholas, Natascha, Mattias, Mathias, Daniel, Annika, Sara, Inga-Lill, Kina, Fereshteh, Abboud, Lubna, Rebecca, Carolina, Emma, Maria, Sabina, Krister, Hanna, Monica N, Niklas, Li, Monica L, Monica S, Jinwen, and Lina.

The rest of the people at "floor 3", and at IMM in general, especially Kristian, Katarina, Therese, Åse, Louise, Johan, and Edyta for pleasant company, beer-nights, and "IMM-parties".

Special thanks to my great family: Maria, Kerstin, Eivin, Ragnar P, Emil, Hillevi, Henning, Ann-Sofi, and Ragnar S. You are the best!

Finally, I would like to thank my mother Lilian for all the love and support you have given me over the years. Thanks for everything!

4 REFERENCES

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994a). Dose-response assessment for developmental toxicity II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. *Fundam Appl Toxicol 23*, 487-495.

Allen, B.C., Kavlock, R.J., Kimmel, C.A., Faustman, E.M. (1994b). Dose-response assessment for developmental toxicity III. Statistical models. *Fundam Appl Toxicol 23*, 496-509.

Allen, B.C., Strong, P.L., Price, C.J., Hubbard, S.A., Daston, G.P. (1996). Benchmark dose analysis of developmental toxicity in rats exposed to boric acid. *Fundam Appl Toxicol 32*, 194-204.

Allen, B.C., Gentry, R., Shipp, A., Landingham, C. (1998). Calculation of benchmark doses for reproductive and developmental toxicity observed after exposure to isopropanol. *Regul Toxicol Pharmacol 28*, 38-44.

Barnes, D., Datson, P., Evans, J., Jarabek, A., Kavlock, R., Kimmel, C., Park, C., Spitzer, H. (1995). Benchmark dose workshop: Criteria for use of a benchmark dose to estimate a reference dose. *Regul Toxicol Pharmacol 21*, 296-306.

Barton, H.A, and Das, S. (1996). Alternatives for a risk assessment on chronic noncancer effects from oral exposure to trichloroethylene. *Regul Toxicol Pharmacol 24*, 269-285.

Barton, H.A., Andersen, M.E., Allen, B.C. (1998). Dose-response characteristics of uterine responses in rats exposed to estrogen agonists. *Regul Toxicol Pharmacol 28*, 133-149.

Birnbaum, L.S., Harris, M.W., Barnhart, E.R., Morrissey, R.E. (1987a). Teratogenicity of three polychlorinated dibenzofurans in c57bl/6n mice. *Toxicol Appl Pharmacol 90*, 206-216.

Birnbaum, L.S., Harris, M.W., Crawford, D.D., Morrissey, R.E. (1987b). Teratogenic effects of polychlorinated dibenzofurans in combination in c57bl/6n mice. *Toxicol Appl Pharmacol 91*, 246-255.

Birnbaum, L.S., Harris, M.W., Stocking, L.M., Clark, A.M., Morrissey, R.E. (1989). Retinoic acid and 2,3,7,8-tetrachlorodibenzo-p-dioxin selectively enhance teratogenesis in c57bl/6n mice. *Toxicol Appl Pharmacol 98*, 487-500.

Birnbaum, L.S., Morrissey, R.E., Harris, M.W. (1991). Teratogenic effects of 2,3,7,8-terabromodibenzo-p-dioxin and three polybrominated dibenzofurans in c57bl/6n mice. *Toxicol Appl Pharmacol 107*, 141-152.

Budtz-Jorgensen, E., Grandjean, P., Kieding, N., White, R.F., Weihe, P. (2000). Benchmark dose calculations of methylmercury-associated neurobehavioral deficits. *Toxicol Lett 112-113*, 193-199.

Burnham, K.P., and Anderson, D.R. (1998). Model selection and inference: a practical information-theoretic approach. Springer, New York.

Casarett, L. J., and Doull, J. (1996). Casarett and Doull's toxicology: The basic science of poisons. Klaassen, C. D., ed. 5th ed. McGraw-Hill Companies, Inc.

Clewell, H.J., Lawrence, G.A., Calne, D.B., Crump, K.S. (2003). Determination of an occupational exposure guideline for manganese using the benchmark method. *Risk Anal 23*, 1031-1046.

Cornish-Bowden, A. (1995). Fundamentals of enzyme kinetics. Portland Press Ltd, London.

Crump, K. (1984). A new method for determining allowable daily intakes. *Fundam Appl Toxicol 4*, 854-871.

Crump, K. (1995). Calculation of benchmark doses from continuous data. *Risk Anal 15*, 79-89.

Crump, K., Viren, J., Silvers, A., Clewell, H., Gearhart, J., Shipp, A. (1995). Reanalysis of dose-response data from the Iraqi methylmercury poisoning episode. *Risk Anal 15*, 523-532.

Crump, K., Kjellstrom, T., Shipp, A., Silvers, A., Stewart, A. (1998). Influence of prenatal mercury exposure upon scholastic and psychological test performance: benchmark analysis of a New Zeeland cohort. *Risk Anal 18*, 701-713.

Crump, K., Van Landingham, C., Shamlaye, C., Cox, C., Davidson, P., Myers, G., Clarkson, W. (2000). Benchmark concentrations for methylmercury obtained from the Seychelles child development study. *Environ Health Perspect 108*, 257-263.

Crump, K. (2002). Critical issues in benchmark calculations from continuous data. *Crit Rew Toxicol 32*, 133-153.

Dekkers, S., de Heer, C., Rennen, M.A. (2001). Critical effect sizes in toxicological risk assessment: a comprehensive and critical evaluation. *Environ Toxicol Pharmacol 10*, 33-52.

Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A.G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R., Younes, M. (2002). Hazard characterisation of chemicals in food and diet: dose response, mechanisms and extrapolation issues. *Food Chem Toxicol 40*, 237-282.

Edler, L., and Kopp-Schneider, A. (1998). Statistical models for low dose exposure. *Mutat Res 405*, 227-236.

Edler, L., Poirier, K., Dourson, M., Kleiner, J., Mileson, B., Nordmann, H., Renwick, A., Slob, W., Walton, K., Wurtzen, G. (2002). Mathematical modeling and quantitative methods. *Food Chem Toxicol 40*, 283-326.

Falk Filipsson, A., Sand, S., Nilsson, J., Victorin, K. (2003). The benchmark dose method – review of available models, and recommendations for application in heath risk assessment. *Crit Rew Toxicol 33*, 505-542.

Falk Filipsson, A., and Victorin, K. (2003). Comparison of available benchmark dose softwares and models using trichloroethylene as a model substance. *Regul Toxicol Pharmacol 37*, 343-355.

Faustman, E.M., Allen, B.C., Kavlock, R.J., Kimmel, C.A. (1994). Dose-response assessment for developmental toxicity I. Characterization of database and determination of no observed adverse effect levels. *Fundam Appl Toxicol 23*, 478-486.

Fletcher, N., Giese, N., Schmidt, C., Stern, N., Lind, P.M., Viluksela, M., Tuomisto, J.T., Tuomisto, J., Nau, H., Håkansson, H. (2005). Altered retinoid metabolism in female Long-Evans and Han/Wistar rats following long-term 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)-treatment. *Toxicol Sci 86*, 264-272.

Fowles, J.R., Alexeeff, G.V., Dodge, D. (1999). The use of benchmark dose methodology with acute inhalation lethality data. *Regul Toxicol Pharmacol 29*, 262-278.

Gaylor, D.W., and Slikker, W. (1990). Risk assessment for neurotoxic effects. *Neurotoxicology 11*, 211-218.

Gaylor, D.W., and Slikker, W. (1994). Modeling for risk assessment of neurotoxic effects. *Risk Anal 14*, 333-338.

Gaylor, D.W. (1996). Quantalization of continuous data for benchmark dose estimation. *Regul Toxicol Pharmacol 24*, 246-250.

Gaylor, D.W., and Chen, J. (1996). Precision of benchmark dose estimates for continuous (nonquantal) measurements of toxic effects. *Regul Toxicol Pharmacol 24*, 19-23.

Gaylor, D., Ryan, L., Krewski, D., Zhu Y. (1998). Procedures for calculating benchmark doses for health risk assessment. *Regul Toxicol Pharmacol 28*, 150-164.

Gaylor, D.W., and Aylward, L.L. (2004). An evaluation of benchmark dose methodology for non-cancer continuous-data health effects in animals due to exposures to dioxin (TCDD). *Regul Toxicol Pharmacol 40*, 9-17.

Gaylor, D.W., and Slikker, W. (2004). Role of the standard deviation in the estimation of benchmark doses with continuous data. *Risk Anal 6*, 1683-1687.

Gephart, L.A., Salminen, W.F., Nicolich, M.J., Pelekis, M. (2001). Evaluation of subchronic toxicity data using the benchmark dose approach. *Regul Toxicol Pharmacol 33*, 37-59.

Giraldo. J., Vivas, N.M., Vila, E., Badia, A. (2002). Assessing the (a)symmetry of concentration-effect curves: empirical versus mechanistic models. *Pharmacol Ther 95*, 21-45.

Haag-Grönlund, M., Fransson-Steen, R., Victorin, K. (1995). Application of the benchmark method to risk assessment of trichloroethene. *Regul Toxicol Pharmacol 21*, 261-269.

Haber, L.T., Allen, B.C., Kimmel, C.A. (1998). Non-cancer risk assessment for nickel compounds: issues associated with dose-response modeling of inhalation and oral exposures. *Toxicol Sci 43*, 213-229.

Jacobson, J.L., Janisse, J., Banerjee, M., Jester, J., Jacobson, S.W., Ager, J. (2002). A benchmark dose analysis of prenatal exposure to polychlorinated biphenyls. *Environ Health Perspect 110*, 393-398.

Jämsä, T., Viluksela, M., Tuomisto, J.T., Tuomisto, J., Tuukkanen, J. (2001). Effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on bone in two rat strains with different aryl hydrocarbon receptor structures. *J Bone Miner Res 16*, 1812-1820.

Kalliomaa, K., Haag-Grönlund, M., Victorin, K. (1998). A new model function for continuous data sets in health risk assessment of chemicals using the benchmark dose concept. *Regul Toxicol Pharmacol 27*, 98-107.

Kavlock, R.J., Allen, B.C., Faustman, E.M., Kimmel, C.A. (1995). Dose-response assessments for developmental toxicity IV. Benchmark doses for fetal weight changes. *Fundam Appl Toxicol 26*, 211-222.

Kavlock, R.J., Schmid, J.E., Setzer, R.W. (1996). A simulation study of the influence of study design on the estimation of benchmark doses for developmental toxicity. *Risk Anal 3*, 399-410.

Kim, A.H., Kohn, M.C., Portier, C.J., Walker, N.J. (2002). Impact of physiologically based pharmacokinetic modeling on benchmark dose calculations for TCDD-induced biochemical responses. *Regul Toxicol Pharmacol 36*, 287-296.

Kimmel, C.A., and Gaylor, D. (1988). Issues in qualitative and quantitative risk analysis for developmental toxicology. *Risk Anal 8*, 15-21.

Kodell, R.L., Howe, R.B., Chen, J.J., Gaylor, D.W. (1991). Mathematical modeling of reproductive and developmental toxic effects for quantitative risk assessment. *Risk Anal 11*, 583-590.

Kodell, R.L., and West, R.W. (1993). Upper confidence limits on excess risk for quantitative responses. *Risk Anal 13*, 177-182.

Kodell, R.L., Chen, J.J., Gaylor, D.W. (1995). Neurotoxicity modeling for risk assessment. *Regul Toxicol Pharmacol 22*, 24-29.

Krewski, D., and Van Ryzin, J. (1981). Dose response models for quantal response toxicity data. In *Statistics and Related Topics* (M. Csorgo, D. Dawson, J. N. K. Rao, and E. Saleh, Eds.), pp. 201-231. North-Holland, Amsterdam.

Krewski, D., and Zhu, Y. (1994). Applications of multinomial dose-response models in developmental toxicity risk assessment. *Risk Anal 14*, 613-627.

Krewski, D., and Zhu, Y. (1995). A simple data transformation for estimating benchmark doses in developmental toxicity experiments. *Risk Anal 15*, 29-39.

Krewski, D., Smythe, R., Fung, K.Y. (2002). Optimal designs for estimating the effective dose in developmental toxicity experiments. *Risk Anal 22*, 1195-1205.

Kupper, L.L., Portier, C., Hogan, M.D., Yamamoto, E. (1986). The impact of litter effects on dose-response modeling in teratology. *Biometrics 42*, 85-98.

Leisenring, W., and Ryan, L. (1992). Statistical properties of the NOAEL. *Regul Toxicol Pharmacol 15*, 161-171.

Mayura, K., Spainhour, C.B., Howie, L., Safe, S., Phillips, T.D. (1993). Teratogenicity and immunotoxicity of 3,3',4,4',5-Pentachlorobiphenyl in c57bl/6 mice. *Toxicology 77*, 123-131.

Murrell, J.A., Portier, C.J., Morris, R.W. (1998). Characterizing dose-response I: Critical assessment of the benchmark dose concept. *Risk Anal 18*, 13-26.

Peto, R., Gray, R., Brantom, P., Grasso, P. (1991). Effects on 4080 rats of chronic ingestion of *N*-nitrosodiethylamine or *N*-nitrosodimethylamine: a detailed dose-response study. *Cancer Res 51*, 6415-6451.

Rai, K., and Van Ryzin, J. (1985). A dose-response model for teratological experiments involving quantal responses. *Biometrics 41*, 1-9.

Rice, D.C. (2004). The US EPA reference dose for methylmercury: sources of uncertainty. *Environ Res 95*, 406-413.

Richards, F. J. (1959). A flexible growth function for empirical use. *J Exp Botany 10*, 290-300.

Setzer, R.W., and Kimmel, C.A. (2003). Use of NOAEL, benchmark dose, and other models for human risk assessment of hormonally active substances. *Pure Appl Chem 75*, 2151-2158.

Slikker, W., Scallet, A., Gaylor, D.W. (1998). Biologically-based dose-response model for neurotoxicity risk assessment. *Toxicol Lett 102-103*, 429-433.

Slikker, W., Andersen, M.E., Bogdanffy, M.S., Bus, J.S., Cohen, S.D., Conolly, R.B., David, R.M., Doerrer. N.G., Dorman, D.C., Gaylor, D.W., Hattis, D., Rogers, J.M., Setzer, R.W., Swenberg, J.A., Wallace, K. (2004a). Dose-dependent transitions in mechanisms of toxicity. *Toxicol Appl Pharmacol 201*, 203-225.

Slikker, W., Andersen, M.E., Bogdanffy, M.S., Bus, J.S., Cohen, S.D., Conolly, R.B., David, R.M., Doerrer. N.G., Dorman, D.C., Gaylor, D.W., Hattis, D., Rogers, J.M., Setzer, R.W., Swenberg, J.A., Wallace, K. (2004b). Dose-dependent transitions in mechanisms of toxicity: case studies. *Toxicol Appl Pharmacol 201*, 226-294.

Slob, W., and Pieters, M.N. (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 Modeling of Continuous Endpoints. *Toxicol Sci 66*, 298-312.

Slob, W., Moerbeek, M., Rauniomaa, E., Piersma, A.H. (2005). A statistical evaluation of toxicity study designs for the estimation of the benchmark dose in continuous endpoints. *Toxicol Sci 84*, 167-185.

Starr, T.B., and Festa, J.L. (2003). A proposed inhalation reference concentration for methanol. *Regul Toxicol Pharmacol 38*, 224-231.

Stiteler, W.M., and Swartout, J. (1991). Some statistical issues relating to the evaluations of risk levels at hazardous waste sites. Proceedings of the section on statistics and the environment, ASA, 39-43.

Toyoshiba, H., Walker, N.J., Bailer, A.J., Portier, C.J. (2004). Evaluation of toxic equivalency factors for induction of cytochromes P450 CYP1A1 and CYP1A2 enzyme activity by dioxin-like compounds. *Toxicol Appl Pharmacol 194*, 156-168.

U.S. Environmental Protection Agency (USEPA). (1995). The use of the benchmark dose (BMD) approach in health risk assessment. Final report. EPA/630/R-94/007. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency (USEPA). (2000). Benchmark dose technical guidance document (BMDTG) (2000). External review draft. EPA/630/R-00/001. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency (USEPA). (2005a). Guidelines for carcinogen risk assessment. Final report. EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency (USEPA). (2005b). The benchmark dose software (BMDS), version 1.3.2. Available at: http://cfpub.epa.gov/ncea/.

Viluksela, M., Bager, Y., Tuomisto, J.T., Scheu, G., Unkila, M., Pohjanvirta, R., Flodström, S., Kosma, V., Mäki-Paakkanen, J., Vartiainen, T., Klimm, C., Schramm, K., Wärngård, L., Tuomisto, J. (2000). Liver tumor-promoting activity of 2,3,7,8-tetrachlodibenzo-p-dioxin (TCDD) in TCDD-sensitive and TCDD-resistant rat strains. *Cancer Res 60*, 6911-6920.

Weber, H., Harris, M.W., Haseman, J.K., Birnbaum, L.S. (1985). Teratogenic potency of TCDD, TCDF and TCDD-TCDF combinations in c57bl/6n mice. *Toxicol Lett 26*, 159-167.

West, R.W., and Kodell, R.L. (1993). Statistical methods of risk assessment for continuous variables. *Commun Statist – Theory Meth 22*, 3363-3376.

West, R.W., and Kodell, R.L. (1999). A comparison of methods of benchmark-dose estimation for continuous response data. *Risk Anal 19*, 453-459.

World Health Organization (WHO). (1999). Principles for the assessment of risks to human health from exposure to chemicals. Environmental Health Criteria 210, Word Health Organization, Geneva.

Zhou, T., Ross, D.G., DeVito, M.J., Crofton, K.M. (2001). Effects of short-term in vivo exposure to polybrominated diphenyl ethers on thyroid hormones and hepatic enzyme activities in weanling rats. *Toxicol Sci 61*, 76-82.

Zhou, T., Taylor, M.M., DeVito, M.J., Crofton, K.M. (2002). Developmental exposure to brominated diphenyl ethers results in thyroid hormone disruption. *Toxicol Sci 66*, 105-116.

Zhu, Y., Jia, Z., Wang, W., Gift, J.S., Moser, V.C., Pierre-Louis, B.J. (2005). Analyses of neurobehavioral screening data: Benchmark dose estimation. *Regul Toxicol Pharmacol 42*, 190-201.