

 Open access • Journal Article • DOI:10.1111/J.0006-341X.2001.00698.X

Benchmark Dose Calculation from Epidemiological Data — [Source link](#)

Esben Budtz-Jørgensen, Niels Keiding, Philippe Grandjean

Institutions: University of Copenhagen, University of Southern Denmark

Published on: 01 Sep 2001 - Biometrics (Biometrics)

Topics: Environmental exposure and Benchmark (computing)

Related papers:

- [A New Method for Determining Allowable Daily Intakes](#)
- [Calculation of Benchmark Doses from Continuous Data](#)
- [Cognitive Deficit in 7-year-old Children With Prenatal Exposure to Methylmercury](#)
- [Critical issues in benchmark calculations from continuous data.](#)
- [Benchmark concentrations for methylmercury obtained from the Seychelles Child Development Study.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/benchmark-dose-calculation-from-epidemiological-data-23ypc649r5>

Benchmark dose calculation from epidemiological data

Budtz-Jørgensen, E.; Keiding, N.; Grandjean, P.

Published in:
Biometrics

DOI:
10.1111/j.0006-341x.2001.00698.x

Publication date:
2001

Document version:
Submitted manuscript

Citation for published version (APA):
Budtz-Jørgensen, E., Keiding, N., & Grandjean, P. (2001). Benchmark dose calculation from epidemiological data. *Biometrics*, 57, 698-706. <https://doi.org/10.1111/j.0006-341x.2001.00698.x>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Benchmark Dose Calculation from Epidemiological Data

Esben Budtz-Jørgensen,^{1,*} Niels Keiding,¹ and Philippe Grandjean²

¹Department of Biostatistics, University of Copenhagen
Blegdamsvej 3, DK-2200 Copenhagen N, Denmark.

²Institute of Public Health, University of Southern Denmark
Winslowparken 17, DK-5000 Odense C, Denmark.

**email:* ebj@biostat.ku.dk

SUMMARY. A threshold for dose-dependent toxicity is crucial for standards setting, but may not be possible to specify from empirical studies. Crump (1984) instead proposed to calculate the lower statistical confidence bound of the benchmark dose, which he defined as the dose that causes a small excess risk. This concept has several advantages and has been adopted by regulatory agencies for establishing safe exposure limits for toxic substances such as mercury. We have examined the validity of this method as applied to an epidemiological study of continuous response data associated with mercury exposure. For models that are linear in the parameters we derived an approximative expression for the lower confidence bound of the benchmark dose. We find that the benchmark calculations are highly dependent upon the choice of the dose-effect function and the definition of the benchmark dose. We therefore recommend that several sets of biologically relevant default settings be used to illustrate the effect on the benchmark results and to stimulate research that will guide an a priori choice of proper default settings.

KEY WORDS: Confidence limits; Environmental epidemiology; Exposure standards; Model dependence; Multiple regression.

1 Introduction

When regulatory agencies produce exposure limits, the decisions are based on available documentation on adverse effects of the chemical in question (WHO, 1994). As thresholds may be difficult to derive from empirical studies, the benchmark dose (BMD) (Crump, 1984, 1995) has been defined as the dose of a toxic compound which increases the probability of an abnormal response by a benchmark response (BMR), i.e., from P_0 for an unexposed subject to $P_0 + \text{BMR}$ for a subject at the BMD. The BMDL is a statistical lower confidence limit of the BMD. An advantage of this approach is that it takes into consideration both biological variation and statistical uncertainty.

The benchmark method is applicable to both categorical and continuous exposure data, including epidemiological data that may involve a continuous exposure scale, a graded response parameter, and potential confounding effects of covariates. Most recently, the Reference Dose for methylmercury of the U.S. Environmental Protection Agency (EPA) was based on benchmark calculations, and a National Academy of Sciences (NAS) committee reviewed and approved the methodology (2000).

The benchmark calculations depend on the default settings for several parameters: A 10% BMR has been proposed for animal experiments on developmental toxicity (Allen et al., 1994). However, a BMR of 5% was used by EPA and NAS, because it corresponds to a doubling of the prevalence of a pathological response, P_0 , which was defined as 5%. As dose-response function, the EPA (1997) used the average for the polynomial and Weibull models for dichotomous methylmercury effects, while for continuous responses, the NAS committee (2000) chose a power function.

The present paper provides a systematical statistical discussion of the benchmark approach as applied in environmental epidemiology. As our example, we use data from a large epidemiological study performed on the Faroe Islands to investigate the health effects of prenatal mercury exposure. This study was identified by the NAS committee (2000) as the critical epidemiological study of mercury toxicity. In adapting the original benchmark concept, the method must be extended to allow confounder correction. An approximative expression for the BMDL in linear models is derived, and the model dependence of the benchmark approach is investigated. Approximative confidence limits for the excess risk at a given dose are derived, and BMDL calculation in more complex linear models is discussed in regard to the wider applicability of this approach.

2 The Faroese Mercury Study

Methylmercury is a common contaminant in seafood and freshwater fish. While adverse effects have been unequivocally demonstrated in poisoning incidents, the implications of lower-level exposures in fish eating populations have been controversial (Grandjean, 1999). This issue was therefore explored in a birth cohort of 1022 children from the Faroe Islands. Information about the children's prenatal exposure was obtained by measuring the mercury concentrations in maternal hair at parturition and in cord blood. The latter biomarker was thought to be the best indicator of the amount of the neurotoxicant that had reached the fetal circulation. Because the effects of fetal exposure to methylmercury are persistent, the children underwent a detailed neuropsychological examination, at age 7 years, when advanced neurobehavioral testing would be feasible. We shall use as an example the Boston Naming Test, a cognitive task reflecting language ability. This test was also used by the NAS committee (2000) to calculate benchmark doses. In this test the child is presented with drawings of objects, which the child has to name. As test score we use the number

of objects that the child failed to identify. Because of the large number of drawings and because on average about half of these drawings are identified the distribution of this outcome variable, given important predictors, is approximately normal. In multiple regression analysis Grandjean et al. (1997) estimated that a 10-fold increase in the cord blood mercury concentration causes a test score deficit of about 1.9 points ($p < 0.0001$).

3 The Benchmark Approach for Experimental Data

The benchmark concept was first developed for standardized experimental dichotomous (normal/abnormal) responses (Crump, 1984), and was later extended to also cover continuous response data (Crump, 1995), which we shall consider. Let $Y(d)$ denote the response of a subject at exposure d . Large responses are assumed to be disadvantageous. The definition of the benchmark dose (BMD) is specific to a dose-response model. In this section only models of the following form will be considered

$$Y(d) = f(d) + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$, that is, at a given dose d the response is assumed to be normally distributed with mean given by the dose-response function f and standard deviation σ . The dose-response function is monotone and may depend on known and unknown parameters. For example, Crump (1984, 1995) suggested a family of power functions, the so-called K -power model: $f(d) = \beta_0 + \beta d^K$, where β_0, β and $K \geq 1$ are parameters to be estimated from the data.

To define the BMD it is necessary to specify the abnormal performance. For continuous data, a cut-off level (x_0) can be specified above which all responses are considered abnormal. The probability of an abnormal response in an unexposed subject is then

$$P_0 = P\{Y(0) > x_0\} = 1 - \Phi\left\{\frac{x_0 - f(0)}{\sigma}\right\}, \quad (1)$$

where Φ is the standard cumulative normal distribution function. Rather than specifying x_0 directly one can also specify the percentage of unexposed subjects whose responses are considered abnormal, i.e. P_0 , x_0 can then be calculated by applying equation (1). Often P_0 is set at 5%. The BMD is defined as the dose that results in an increased probability of an abnormal test performance by a benchmark response (BMR), i.e., the BMD satisfies $P\{Y(\text{BMD}) > x_0\} - P_0 = \text{BMR}$. Figure 1 gives a graphical illustration of the BMD definition.

The BMDL is calculated as the statistical 95% lower (one-sided) confidence bound of the BMD. In this way the power of the study is taken into account. The less the precision the lower the BMDL.

The BMR is often set at 5% so that the corresponding BMD will double the risk of an abnormal response, given that P_0 is 5% (NAS, 2000). Everything else being equal, lower BMRs will result in lower BMDs.

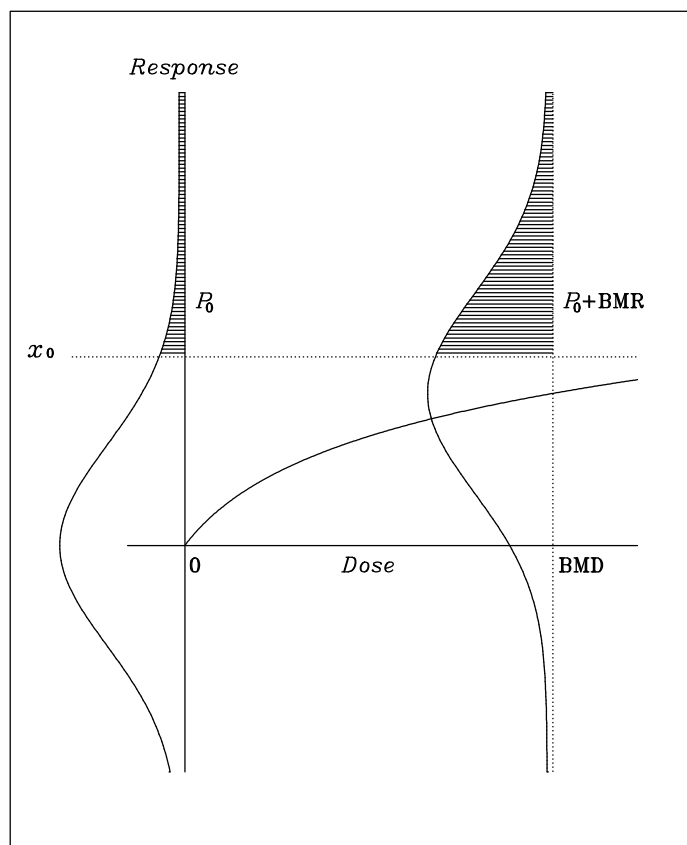


Figure 1: Hypothetical dose-response relation illustrating the concepts of benchmark approach. The dose-response curve indicates that when the dose increases so does the expected response. The distribution of responses in unexposed subjects is shown on the y -axis. Responses above the prespecified x_0 are considered abnormal. The risk of an abnormal response in unexposed subjects is P_0 indicated by the shaded area. At the BMD the response distribution has been translated upward and the risk of an abnormal response has increased to $P_0 + \text{BMR}$. The BMDL is placed somewhere between 0 and the estimated BMD depending on the amount of information in the study.

4 The Benchmark Approach in Observational Studies

Observational studies usually do not include a control group completely free of exposure, the exposures are distributed continuously, and the response variable is influenced by confounders in addition to the exposure of interest. Therefore, let $Y(d, c_1, \dots, c_k)$ denote the response of a subject at exposure d and with confounder values c_1, \dots, c_k . We first assume that the dependence of the response on the exposure and the confounders can be modeled using a multivariate additive model

$$Y(d, c_1, \dots, c_k) = f(d) + h(c_1, c_2, \dots, c_k) + \epsilon, \quad (2)$$

where $\epsilon \sim N(0, \sigma^2)$ and h is any real valued function typically depending on unknown parameters. The question now is how to define the BMD in the presence of the confounders. Still the main idea is that the BMD is the dose that increases the probability of an abnormal response by a prespecified amount (BMR), i.e., $P\{Y(\text{BMD}, c_1, \dots, c_k) > x_0\} - P\{Y(0, c_1, \dots, c_k) > x_0\} = \text{BMR}$. Using the additivity of the model it is easily seen that this statement is equivalent to

$$\Phi^{-1}[1 - P\{Y(0, c_1, \dots, c_k) > x_0\}] - \Phi^{-1}[1 - P\{Y(\text{BMD}, c_1, \dots, c_k) > x_0\}] = \text{BMR} / \sigma. \quad (3)$$

In the extended benchmark method the level of abnormal test performance x_0 is specified such that all unexposed subjects have the same (prespecified) risk P_0 of an abnormal response, i.e. $P\{Y(0, c_1, \dots, c_k) > x_0\} = P_0$ for all c_1, \dots, c_k . Using this definition, the corresponding BMD (and BMDL) will not depend on the confounders, as can be seen from equation (3), and exposure to the BMD causes the risk of an abnormal response to increase from P_0 to $P_0 + \text{BMR}$. An interpretation in terms of expected test performance is given by equation (3): For each subject the BMD increases the expected response by a prespecified multiple $\Omega = \Phi^{-1}(1 - P_0) - \Phi^{-1}(1 - P_0 - \text{BMR})$ of the standard deviation σ of the random response component.

The main change when extending the benchmark method to epidemiological data is therefore to let the definition of an abnormal response depend on the confounders. If for instance P_0 is set at 5%, then the response of a given subject is considered abnormal if it is above the 95th percentile in the distribution of responses from unexposed subjects with the same values on the confounders as the subject at hand. Although not quite explicitly stated, a similar approach seems to have been used by Crump et al. (1998, 2000). In multiple linear regression Bailer et al. (1997) defined the absolute exposure concentration as the dose causing the expected test score to increase Δ response units compared to unexposed subjects with the same confounder profile. This definition corresponds to a benchmark dose with $P_0 = 1 - \Phi(\Delta/\sigma)$ and $\text{BMR} = 0.5 - P_0$.

5 BMDLs in Models Linear in the Parameters

In models which depend linearly on the unknown parameters, it is possible to obtain closed expressions for the BMD as well as approximative expressions for the BMDL. Consider the model

$$Y(d, c_1, \dots, c_k) = \beta_0 + \beta g(d) + \beta_1 c_1 + \dots + \beta_k c_k + \epsilon, \quad (4)$$

where $\epsilon \sim N(0, \sigma^2)$. The function g (which is assumed known) is increasing and $g(0) = 0$. Crump (1995) required that $g'(0) < \infty$ from the point of view that dose-response functions with an infinite slope at zero dose are not biologically plausible.

The BMD is determined by solving equation (3), yielding:

$$\text{BMD} = \begin{cases} \infty & \text{if } \beta \leq 0 \\ g^{-1}\left(\frac{\Omega\sigma}{\beta}\right) & \text{if } \beta > 0 \end{cases}$$

where $\Omega = \Phi^{-1}(1 - P_0) - \Phi^{-1}(1 - P_0 - \text{BMR})$. Note that if $\beta \leq 0$ there is no solution to equation (3) since an increasing dose will not increase the expected response. A natural way to estimate the BMD is to substitute the unknown parameters in the BMD expression by estimates:

$$\widehat{\text{BMD}} = \begin{cases} \infty & \text{if } \hat{\beta} \leq 0 \\ g^{-1}\left(\frac{\Omega\hat{\sigma}}{\hat{\beta}}\right) & \text{if } \hat{\beta} > 0 \end{cases}$$

This estimator is consistent but clearly biased ($E\{\widehat{\text{BMD}}\} = \infty$). However, this bias is of minor concern if (as is usually the case) the main aim of the benchmark analysis is to calculate a lower confidence limit of the BMD.

5.1 BMDL calculation

There are several ways to calculate the confidence limits, in particular Crump (1984, 1995) suggested likelihood-based limits for the BMDL calculation. Here the lower 95% confidence limit is approximated by the 5th percentile in the estimated distribution of $\widehat{\text{BMD}}$. This method is preferred since it yields a closed expression for the BMDL.

The 5th percentile (γ) in the distribution of $\widehat{\text{BMD}}$ is approximated. We assume as above that large values are detrimental and first consider the likely case where $P(\hat{\beta} > 0) > 5\%$. In this case γ is finite and satisfies $P\{g^{-1}(\Omega\hat{\sigma}/\hat{\beta}) \leq \gamma, \hat{\beta} > 0\} = 5\%$ which can easily be rewritten as $P\{\sqrt{v}\hat{\beta}/\hat{\sigma} \leq \Omega\sqrt{v}/g(\gamma)\} = 95\%$, where v is the amount information on the exposure parameter, e.g., for simple linear regression with n observations $v = \sum_{i=1}^n (d_i - \bar{d})^2$.

The random variable $\sqrt{v}\hat{\beta}/\hat{\sigma}$ has a noncentral t -distribution with df degrees of freedom and noncentrality parameter $\sqrt{v}\beta/\sigma$. According to Johnson and Welch (1939), the 95th percentile in this distribution can be approximated by

$$p_{95} = \frac{\sqrt{v}\beta + u_{95}\sqrt{\sigma^2 + (c\beta^2 - u_{95}^2\sigma^2)/2df}}{\sigma(1 - u_{95}^2/2df)}, \quad (5)$$

where $u_{95} \approx 1.645$ is the 95th percentile in the standard normal distribution. Thus $p_{95} \approx \Omega\sqrt{v}/g(\gamma)$ which means that $\gamma \approx g^{-1}(\Omega\sqrt{v}/p_{95})$. In the rare situation of a strongly *beneficial* substance with $\beta < -u_{95}\sigma/\sqrt{v}$, equivalent to $P(\hat{\beta} > 0) < 5\%$, the probability of an finite BMD estimate is below 5% and $\gamma = \infty$.

The lower 95% confidence bound on the BMD used here is the 5th percentile in the estimated distribution of the $\widehat{\text{BMD}}$, which is obtained simply by substituting estimates for the unknown parameters in the expression for γ

$$\text{BMDL} = \begin{cases} \infty & \text{if } t = \sqrt{v}\hat{\beta}/\hat{\sigma} \leq -u_{95} \\ g^{-1}\left\{\frac{\Omega\hat{\sigma}(1-u_{95}^2/2df)}{\hat{\beta}+u_{95}\widehat{\text{SE}}(\hat{\beta})\sqrt{1+(t^2-u_{95}^2)/2df}}\right\} & \text{if } t = \sqrt{v}\hat{\beta}/\hat{\sigma} > -u_{95} \end{cases} \quad (6)$$

where $\widehat{\text{SE}}(\hat{\beta}) = \hat{\sigma}/\sqrt{v}$ is the estimated standard deviation of $\hat{\beta}$.

It is seen that BMDL is easily obtained from standard multiple regression output. Note also that for an increasing number of the degrees of freedom the BMDL as approximated by (6) approaches (from below) the crude approximative confidence limit obtained by assuming that the residual variance is known, i.e. $\text{BMDL} = g^{-1}[\Omega\hat{\sigma}/\{\hat{\beta} + u_{95} \cdot \widehat{\text{SE}}(\hat{\beta})\}]$ if $t = \sqrt{v}\hat{\beta}/\hat{\sigma} > -u_{95}$. Thus, the terms $(1 - u_{95}^2/2df)$ and $\sqrt{1 + (t^2 - u_{95}^2)/2df}$ can be interpreted as correcting for the uncertainty in $\hat{\sigma}$.

The BMDL decreases as a function of the estimated exposure effect ($\hat{\beta}$) and increases as a function of exposure parameter information (v) and BMR. An increase in the random response variation ($\hat{\sigma}^2$) has two opposing effects. On one hand the BMDL is lowered as a result of increased estimation uncertainty. On the other hand the response distribution becomes more dispersed which will lead to a higher $\widehat{\text{BMD}}$ because the same increase in expected test performance will correspond to a smaller increase in the risk of an abnormal response. When $\hat{\beta} > 0$ the latter effect is the stronger so that the BMDL increases as a function of $\hat{\sigma}^2$. Thus, if a strong predictor of the response which is not related to the exposure is excluded from the set of independent variables then a *higher* BMDL is obtained.

5.2 Dichotomous response data

The extended benchmark method cannot be used directly for 0/1-responses if confounding is present. For such data a response is abnormal simply if it is 1, so the parameter x_0 is not

available. This means that the probability of an abnormal response in unexposed subjects will depend on the confounder values and so will the BMD (Bailer et al., 1997). However, for logistic regression models ($\text{logit}[P\{Y(d, c_1, \dots, c_k) = 1\}] = \beta_0 + \beta g(d) + \beta_1 c_1 + \dots + \beta_k c_k$) it seems obvious to define the BMD as the dose resulting in a prespecified proportional increase (BMR) in the *odds* for an abnormal response, i.e.,

$$\frac{P\{Y(\text{BMD}, c_1, \dots, c_k) = 1\} / [1 - P\{Y(\text{BMD}, c_1, \dots, c_k) = 1\}]}{P\{Y(0, c_1, \dots, c_k) = 1\} / [1 - P\{Y(0, c_1, \dots, c_k) = 1\}]} = \text{BMR}. \quad (7)$$

Solving (7) for BMD, yields $\text{BMD} = g^{-1}\{\log(\text{BMR})/\beta\}$ if $\beta > 0$, independent of confounder values. The BMDL is obtained using the approach described above and the asymptotic normality of $\hat{\beta}$, as $\text{BMDL} = g^{-1}[\log(\text{BMR})/\{\hat{\beta} + u_{95}\widehat{\text{SE}}(\hat{\beta})\}]$ if $\hat{\beta} > -u_{95}\widehat{\text{SE}}(\hat{\beta})$. Alho and Valtonen (1995) derived an algorithm giving likelihood based confidence limits on the dose causing a prespecified increase in the linear predictor in generalized linear models with known scale parameter.

6 Model Dependence of BMDL and BMD

To study the model dependence of the benchmark approach BMDL-values are compared for a linear model i.e. $g(d) = d$, and a power model i.e. $g(d) = d^\alpha$, where α is a known positive number. For simplicity it is assumed that no covariates are present and that the exposure variable d follows a logarithmically normal distribution with parameters $[E\{\log(d)\} = \mu, \text{var}\{\log(d)\} = \tau^2]$. These assumptions yield a parametric expression for the exposure parameter information $v = \Sigma_{i=1}^n \{g(d_i) - \bar{g}(\bar{d})\}^2 \approx n \cdot \text{var}\{g(d)\} = n \cdot \{\exp(\alpha^2 \tau^2) - 1\} \cdot \exp(\alpha^2 \tau^2 + 2\alpha\mu)$.

Model dependence is especially a problem if two models describing the data almost equally well yield very different estimates of the parameter of interest. If the linear model and the power model explain about the same percentage of the response variation, the exposure t -statistics ($\sqrt{v}\hat{\beta}/\hat{\sigma}$) in the two models will be approximately equal. Thus, using the crude approximative expression (section 5.1), the ratio between BMDLs from the linear model and the power model becomes

$$\frac{\text{BMDL}_1}{\text{BMDL}_\alpha} \approx \sqrt{\frac{\{\exp(\tau^2) - 1\} \cdot \exp(\tau^2)}{[\{\exp(\alpha^2 \tau^2) - 1\} \cdot \exp(\alpha^2 \tau^2)]^{\frac{1}{\alpha}}}} \cdot \left(\frac{t + u_{95}}{\sqrt{n} \Omega}\right)^{\frac{1}{\alpha} - 1},$$

where t denotes the mutual level of the two test statistics. This ratio is 1 when $\alpha = 1$ and can be shown to decrease monotonically as a function of α unless $\Omega > 5.63 \cdot (t + u_{95}) / \sqrt{n}$. So except when BMR is high, the lower the α the lower the BMDL. Furthermore, it is seen that the BMDL model dependence becomes stronger for more significant dose-response relations and for lower BMRs. As BMR decreases toward zero this ratio increases beyond

all bounds or decreases toward 0 depending on whether α is below or above 1. In Figure 2 the ratio as a function of α is shown for selected values of the other parameters.

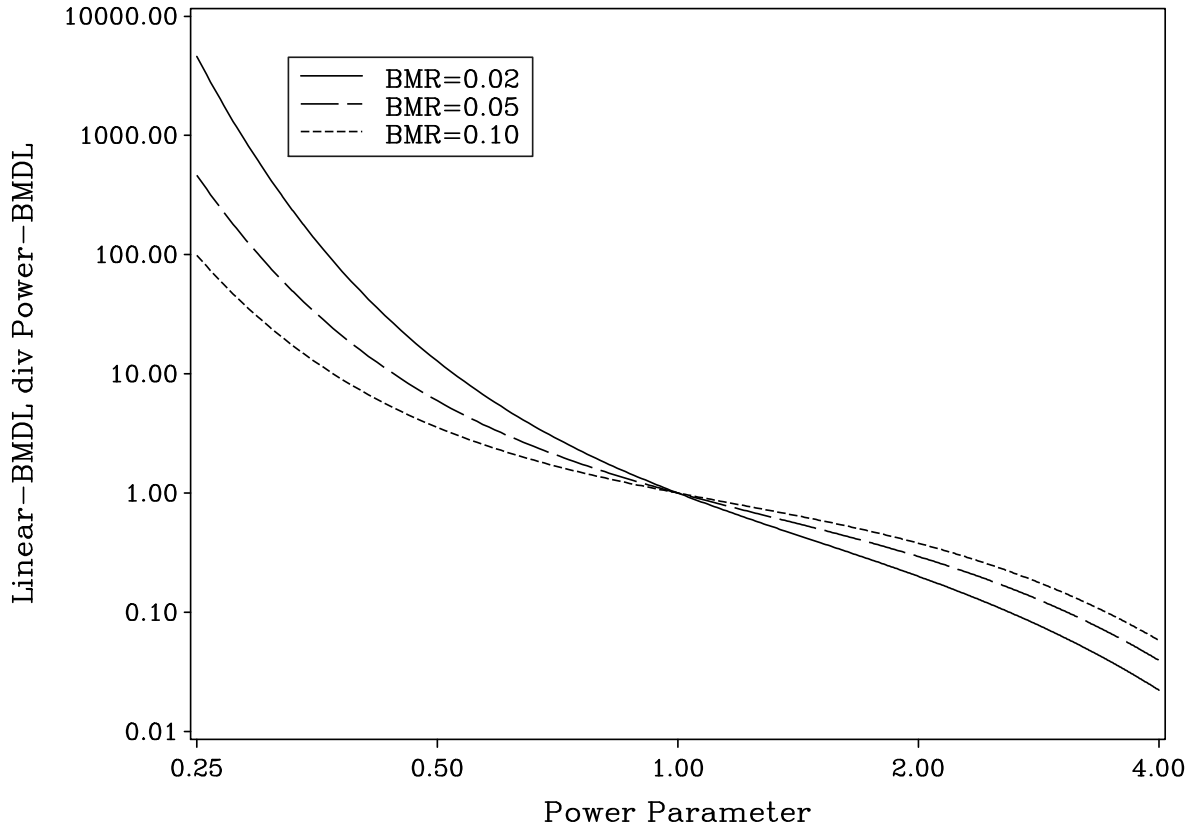


Figure 2: Ratio between the BMDL of the linear dose-response model and the BMDL of a power function model with known power parameter. This ratio is shown as a function of the power parameter for BMR=2, 5 or 10% and $P_0 = 5\%$, $n = 100$, $t = 2.00$, $\tau^2 = 1.00$.

Model dependence is reduced if dose-response functions with infinite slope at zero dose are avoided. Using two different dose-response functions f_1 and f_2 , not necessarily linear in the parameters but with finite positive slope at zero dose, it is easily shown that $\widehat{\text{BMD}}_{f_1}/\widehat{\text{BMD}}_{f_2} \rightarrow f_2'(0)\hat{\sigma}_{f_1}/f_1'(0)\hat{\sigma}_{f_2}$ for $\text{BMR} \rightarrow 0$, where $\hat{\sigma}_{f_1}$ and $\hat{\sigma}_{f_2}$ are the estimated standard deviations of the random response component assuming f_1 and f_2 , respectively. Thus, the ratio between $\widehat{\text{BMD}}$ s estimated in different models stays bounded even as the BMR approaches zero.

7 BMDLs in More Complex Dose-Response Models

Under the model assumptions (2) the *excess risk of an abnormal response at dose d* is

$$\text{ER}(d) = P\{Y(d, c_1, \dots, c_n) > x_0\} - P\{Y(0, c_1, \dots, c_n) > x_0\} \quad (8)$$

$$= 1 - \Phi[\Phi^{-1}(1 - P_0) + \{f(0) - f(d)\}/\sigma] - P_0 \quad (9)$$

It is noticed that this increase in risk is the same for all confounder values. Under the linear models (4) the above expression simplifies to $\text{ER}(d) = 1 - \Phi\{\Phi^{-1}(1 - P_0) - \beta \cdot g(d)/\sigma\} - P_0$, which is consistently estimated by substituting the unknown parameters by their estimates. An upper 95% confidence limit on $\text{ER}(d)$ is obtained by again exploiting that $\sqrt{v}\hat{\beta}/\hat{\sigma}$ has a noncentral t -distribution, yielding $\text{ER}(d)_{95} = 1 - P_0 - \Phi\{\Phi^{-1}(1 - P_0) - \hat{p}_{95}g(d)/\sqrt{v}\}$ where \hat{p}_{95} is obtained from (5) substituting parameters by estimates.

Also notice, that since the risk of an abnormal response at dose d ($P\{d\}$) does not depend on the confounders ordinary dichotomous data dose-response functions for $P(d)$ may be applied by specifying $f(d)$ accordingly: $P(d) = q(d)$ if $f(d) = f(0) + \sigma\Phi^{-1}(1 - P_0) - \sigma\Phi^{-1}\{1 - q(d)\}$. For responses with no confounding this correspondence between dose-response models for continuous and dichotomous data in the benchmark approach was noted by Crump (1995). The present generalization depends on the confounders entering the model in an additive fashion.

In section 5 an approximative expression for BMDLs was obtained in linear models where the dependence of the expected response on the exposure is described by a single term: $\beta \cdot g(d)$. Although these models are quite flexible they may not always suffice. For example, the exposure effect may depend on the level of an effect modifier. Within the set of linear models more complex dose-response relations can be modeled by increasing the number of terms depending on the dose. Ordinary polynomials and fractional polynomials (Royston and Altman, 1994) are well known examples of functions having multiple terms. In these models it may not be possible to derive a closed expression for the BMD. However, upper confidence limits on the excess risk at exposure d (9) can be determined exactly as described above since $\{\hat{f}(0) - \hat{f}(d)\}/\hat{\sigma}$ has a (scaled) noncentral t -distribution. The BMDL can then be determined by solving the equation $\text{ER}(d)_{95} = \text{BMR}$ in d . In the presence of effect modification both the BMD and the BMDL will depend on level of the effect modifier, reflecting that not all subjects are equally sensitive.

Furthermore, the benchmark approach can easily be extended to models where the residual variance is proportional to a known increasing function of the exposure dose. In linear models it will be possible to derive expressions for BMDs and BMDLs using the methods described above.

8 Application of Benchmark Calculations

Previous analyses of mercury toxicity used data from studies carried out in Iraq (Crump et al., 1995), New Zealand (Crump et al., 1998) and the Seychelles (Crump et al., 2000). In these studies the K -power model was used for the dependence of a child's test score on the mercury dose, yielding only convex dose-response functions. To make our results as comparable as possible to those results, we have used the same model. However, as the dose-effect could also be concave within some dose ranges other functions were also appropriate. To illustrate model dependence, $\widehat{\text{BMD}}$ s and BMDLs were therefore also calculated for a linear model ($g\{d\} = d$), a square root model ($g\{d\} = \sqrt{d+1} - 1$) and a logarithmic model ($g\{d\} = \log\{d+1\}$). None of the dose-response curves considered has an infinite slope at zero dose. Confounder adjustment is carried out as indicated above using the same set of confounders which was identified in Grandjean et al. (1997).

In the K -power model the BMD is given as $(\Omega\sigma/\beta)^{1/K}$, which is estimated by substituting parameters by estimates (obtained using the SAS procedure NLIN). With unknown K this model is not linear in the parameters, so the theory described in section 5 does not apply. Instead, the BMDL is calculated using the parametric bootstrap method (Efron and Tibshirani, 1993). Again, the lower confidence limit of the BMD is approximated by the 5th percentile in the estimated distribution of $\widehat{\text{BMD}}$, but here this percentile is determined by simulating 2000 new data sets from the estimated distribution of the original data. Each data set has covariate values identical to the values of the original data whereas the response is given by the function of the covariates estimated in the original data plus a random normal error with variance $\hat{\sigma}^2$. After each simulation, the BMD is estimated, and the BMDL is then determined as the 5th percentile in the empirical distribution of the estimates.

Table 1

$\widehat{\text{BMD}}$ s and BMDLs ($\mu\text{g/l}$) calculated for the cord blood mercury concentration for different BMRs and $P_0 = 5\%$. Further, $-2\cdot\log(L)$ indicates minus twice the log of the likelihood function, while t is the usual test statistic for no effect of the exposure.

Model	$-2\cdot\log(L)$	t	BMR=0.02		BMR=0.05		BMR=0.10	
			$\widehat{\text{BMD}}$	BMDL	$\widehat{\text{BMD}}$	BMDL	$\widehat{\text{BMD}}$	BMDL
K -power ^a	7793.72	-	39.55	28.73	84.98	61.22	142.32	102.22
Linear	7793.72	3.49	39.55	26.81	84.98	57.61	142.32	96.49
Square Root	7790.84	3.89	11.55	6.73	40.78	22.30	102.03	53.91
Logarithmic	7790.42	3.94	1.55	0.93	6.46	3.11	27.94	9.67

^a $\widehat{K} = 1$.

Table 2

$\widehat{\text{BMD}}$ s and BMDL s ($\mu\text{g/g}$) calculated for the maternal hair mercury concentration for different BMR s and $P_0 = 5\%$. Further, $-2 \cdot \log(L)$ indicates minus twice the log of the likelihood function, while t is the usual test statistic for no effect of the exposure.

Model	$-2 \cdot \log(L)$	t	$\widehat{\text{BMR}}=0.02$		$\widehat{\text{BMR}}=0.05$		$\widehat{\text{BMR}}=0.10$	
			$\widehat{\text{BMD}}$	BMDL	$\widehat{\text{BMD}}$	BMDL	$\widehat{\text{BMD}}$	BMDL
K -power ^a	8021.66	-	7.08	4.79	15.22	10.02	25.49	16.69
Linear	8021.66	2.90	7.08	4.51	15.22	9.69	25.49	16.22
Square Root	8021.71	2.89	4.03	2.20	12.48	6.28	28.95	13.77
Logarithmic	8022.58	2.73	1.83	0.91	8.35	3.03	41.27	9.30

^a $\widehat{K} = 1$.

Tables 1 and 2 give estimated $\widehat{\text{BMD}}$ s and BMDL s for mercury exposure expressed as the cord blood concentration and the maternal hair concentration, respectively.

For both exposure variables, the power parameter in the K -power model is estimated to 1, thus yielding a linear model. This means that the $\widehat{\text{BMD}}$ of the K -power model is equal to that of the linear model, but the BMDL s are slightly different because of the additional parameter estimated in the K -power model and because of simulation uncertainty.

The fits of the linear, the square root and the logarithmic models were compared in pairs. Each model in a given pair was tested against an expanded model that included both models. A high p -value in this test indicates that the model at hand does not fit the data significantly worse than the expanded model. For all pairs, the expanded model fitted the data only slightly better than the best of the models in the pair. This means that the models can be compared simply by calculating their $-2 \log(L)$ -difference and then judging this value from a χ_1^2 -table. For the cord blood mercury, there is no difference between the fit of the logarithmic and the square root model, but both of these models are marginally significantly better than the linear model ($p = 6\% - 8\%$). For the hair mercury results, the data do not allow a judgment as to which model is the best. The fit of the K -power model cannot be compared to the rest of the models by formal asymptotic likelihood testing, because this model has estimates on the boundary of the parameter space ($\widehat{K} = 1$). However, as the K -power model gives the same fit as the linear model but has one more (redundant) parameter, at best this model can be considered only as good as the linear model.

The BMDL s for the cord blood mercury exhibit a substantial variation across the models, with the logarithmic model giving the lowest results. For $\text{BMR}=2\%$ the BMDL of the logarithmic model and the linear model differ by a factor of almost 30, while for $\text{BMR}=10\%$ this factor has dropped to about 10. Also, the results for the square root model are

considerably higher than for the logarithmic model, although the difference in fit between these two models was far from significant. Since the model uncertainty is larger and the mercury effect weaker when using the maternal hair concentration as the exposure variable it is no surprise that the observed model dependence of the BMDL is weaker. Again, the logarithmic model yields the lowest results, while the linear model yields the highest. However, the ratio of the results is now only about 5 for BMR=2% and 1.7 for BMR=10%.

The square root and the logarithmic models obviously depend on the choice of scale of the dose, which defines the meaning of the constant 1 in these models. We find (data not shown) that the lower the constant the lower the BMDL. However, the BMDLs of a square root model without adding a constant ($g\{d\} = \sqrt{d}$) were still not as low as the results of the logarithmic model (with constant 1).

In Figure 3, the model dependence for the cord blood concentration is illustrated graphically. For each model the estimated expected excess response due to mercury exposure is shown as a function of the mercury dose ($d \rightarrow \hat{f}\{d\} - \hat{f}\{0\}$). The BMD is the dose which causes a certain level of excess response, i.e: $f(\text{BMD}) - f(0) = \Omega\sigma$. Since the estimates of σ are approximately the same in all the models, the $\widehat{\text{BMD}}$ s can (approximately) be found as the intersections between the excess response curves and the same horizontal line. The curves are clearly different: although 95% of the observations have a mercury concentration between 5 and 80 $\mu\text{g}/\text{l}$, the logarithmic curve is far above the linear dose-response curve across this range. It may seem strange that there is not enough information in the data from almost 1000 children to determine which of these curves is the best. However, the curves shown are not the estimated dose-response functions ($d \rightarrow \hat{f}\{d\}$) which are much closer across the actual dose range and intersect twice. The excess response curves differ more because of the large differences in the estimated response level of unexposed children ($\hat{f}\{0\}$), determined by extrapolation. Compared to a convex curve (the linear) a concave dose-response curve (the logarithmic) will give a lower unexposed response level, a higher slope at low doses and a lower slope at high doses. Thus, at low BMRs the corresponding concave excess response curve will yield lower $\widehat{\text{BMD}}$ s and BMDLs. At a certain BMR the excess response curves meet, yielding identical $\widehat{\text{BMD}}$ s. However, the BMDL of the concave curve will still be lower because the concave curve is not as steep as convex one when the two curves meet. Thus, the point of intersection with the horizontal line is more uncertain for the concave curve, thereby giving wider BMD-confidence bands. An example of this is seen in Table 2: For BMR=10% the logarithmic $\widehat{\text{BMD}}$ is higher than the linear $\widehat{\text{BMD}}$ while the opposite relation is seen for the BMDLs.

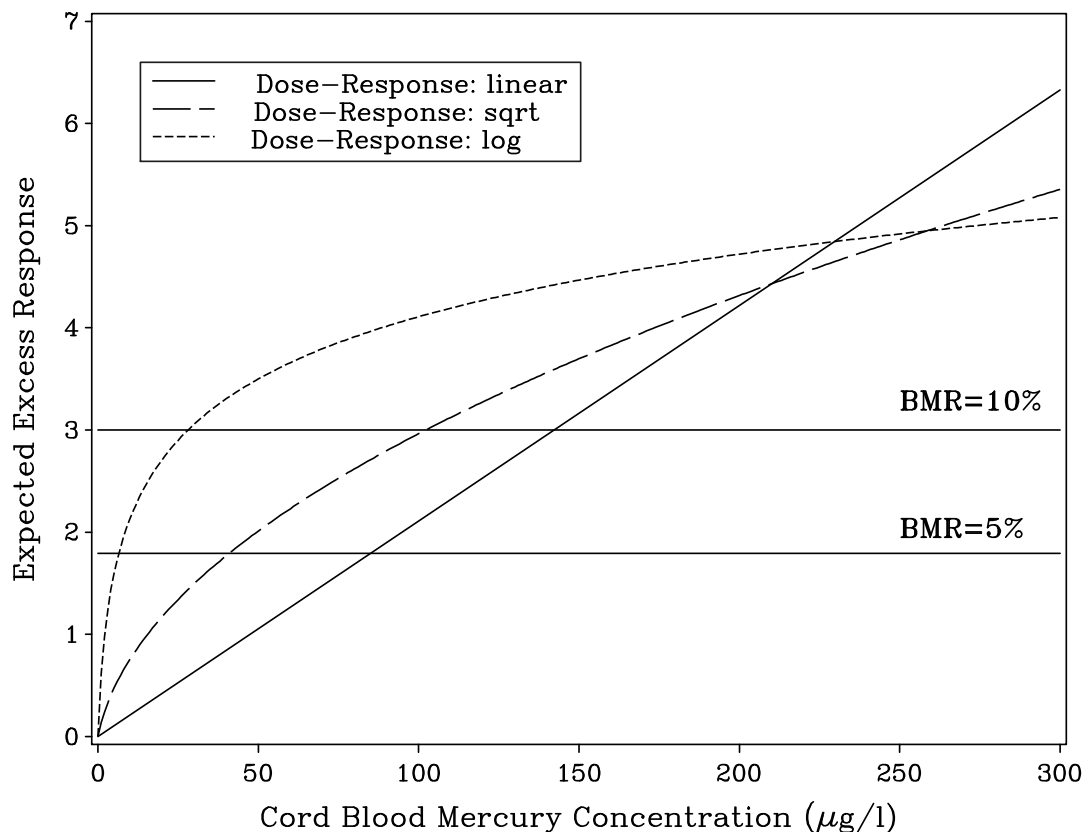


Figure 3: For each model the estimated expected excess response due to mercury exposure is shown as a function of the mercury dose ($d \rightarrow \hat{f}(d) - \hat{f}(0)$). The BMD is the dose which causes a certain level of excess response given as a multiple of the standard deviation of the random response component, i.e: $f(\text{BMD}) - f(0) = \Omega\sigma$. Since the estimates of σ are approximately the same in all the models, the BMDs can (approximately) be found as the intersections between the excess response curves and the same horizontal line. Lines corresponding to $P_0 = 5\%$ and BMR=5%, 10% are indicated on the graph.

8.1 Validity of Confidence Limits

Except for the K -power model, all BMDL-values are calculated using the formula (6). The validity of these confidence limits was investigated by simulations. For each model, one million data sets was simulated from the estimated distribution of the original data. After each simulation the BMDL was calculated and then compared to the known BMD. Table 3 shows the empirical coverage probability. For all models and both exposure variables,

the coverage probability is very close to the nominal value of 95%.

Table 3

Empirical coverage probability in percent of BMDL in 1,000,000 simulations from the estimated distribution of the response variable in the Faroese data. The standard deviation of the estimated coverage probability is 0.022%.

	Dose-Response Model		
	Linear	Square Root	Logarithmic
Blood Mercury	95.0426	95.1036	95.0404
Hair Mercury	94.9921	95.0493	95.0089

For each of the models the set of observations satisfying that the BMDL is less than the BMD is independent of the prespecified parameters P_0 and BMR. This means that the coverage probability of the BMDL is the same for all combinations of P_0 and BMR.

9 Discussion

Statistical methods for determination of safe dose levels have largely been aimed at being used for dichotomous responses from animal experiments (WHO, 1994). For more complex epidemiological data the extended benchmark method can be applied with certain reservations.

The closed approximative expression for the BMDL will facilitate future calculations and improve understanding of the benchmark approach. Kodell and West (1993) derived a related method for BMDL calculation in an experimental setting also based on the noncentral t -distribution. However, this method is iterative and does not provide a closed expression for the BMDL. Furthermore, Johnson and Welch (1939) showed that the approximation to the noncentral t -distribution used here is superior to the one used by Kodell and West.

A priori specification of P_0 and BMR is a fundamental difficulty in the benchmark approach. Ideally, these specifications should be based on biological considerations. However, often the biological understanding of the mechanism under consideration is inadequate and instead it seems that a P_0 of 5% and a BMR of 5% or 10% have somewhat arbitrarily been chosen as the default values (EPA, 1997; Crump et al., 1998, 2000; NAS, 2000). The results of the Faroese study clearly underline that the choice of BMR is critical: both for the square root model and the logarithmic model did the cord blood results increase by a factor of about 10 when the BMR was changed from 2% to 10% and P_0 was fixed at 5%. The less critical sensitivity of the BMDL to specification of P_0 has been investigated in additional calculations (data not shown). When P_0 was increased from 5% to 16%, cord

blood BMDLs decreased by 40% to 70%. When deciding whether a given set of values for P_0 and BMR corresponds to a biologically significant effect for a given endpoint it might be helpful to calculate $\Omega \cdot \hat{\sigma}$, which is the expected excess response of a subject at the BMD. Figure 3 shows that a P_0 and a BMR of 5% corresponds to an expected excess response of about 1.8 points on the Boston Naming Test, while the expected score is increased by approximately 3 points at the BMD if the BMR is changed to 10%. The implications of such effects should be considered from a neurological viewpoint.

In the extended method, the cut-off x_0 separating an abnormal response from a normal one is not constant but is calculated such that, for all unexposed subjects, the risk of an abnormal response is equal to a prespecified amount P_0 . In accordance with the original method, the BMD is defined as the dose which increases the risk from P_0 to $P_0 + \text{BMR}$. In addition to being mathematically convenient in that the result does not depend on the confounders, this extension of the benchmark method seems natural because it corresponds to doing an ordinary benchmark calculation for each combination of the confounder values (with the same prespecified P_0 for each combination). An alternative approach would have been to fix only one (confounder independent) cut-off value and to define the BMD as the dose causing a certain increased risk of exceeding this level. It is easily seen from (3) that even in a model assuming no effect modification this approach would yield covariate-dependent BMDs. Thus, this covariate dependence is solely a consequence of not taking into account differences in response level between unexposed subjects with different covariate profiles when defining the abnormal response level.

However, application of the benchmark method to epidemiological data may be problematic, because the definition of BMD is dependent on the response level of unexposed subjects. When an unexposed control group is not available, the unexposed response level is determined by extrapolation, thereby making the benchmark result sensitive to the selection of dose-response model. In addition, many response variables are noisy. In the Faroese data, about 80% of the variation in the test scores on the Boston Naming Test could not be explained by known predictors. A consequence of this noise is a limited power to distinguish between dose-response functions that have relatively different (concave/convex) curve shapes in the observed dose-range (and therefore might be very different at zero dose). Thus, even though the models considered may yield very different BMDLs, they may fit the data almost equally well. In the relatively large Faroese study, the model dependence had serious consequences: BMDLs from the linear and the logarithmic dose-response models differed by a factor between 10 to 30, depending on the selected BMR. For smaller studies, the model dependence will probably be even greater.

On the subject of extrapolation and model dependence, the standard advice is (Crump, 1995): The BMR is typically set at the lower end of the range that can be detected experimentally, in order to avoid uncertainties associated with low-dose extrapolation using models that may not reflect biological realities. This statement assumes the existence of a

dose range in which the estimated excess response curves ($d \rightarrow \hat{f}\{d\} - \hat{f}\{0\}$) run closely together. Such a range may exist for the dose-response curves ($d \rightarrow \hat{f}\{d\}$). However, as a result of the translation given by the differences in the estimated unexposed response level (as determined by extrapolation), the excess response curves may be very different in the whole dose-range, yielding highly model-dependent results.

The calculations presented in this paper emphasize that the benchmark approach is useful to derive guidance from epidemiological data for the purpose of defining safe exposures to toxic compounds, such as mercury. However, the uncertainties inherent in such data may have severe implications for some calculations of benchmark doses. Benchmark calculations are no more precise than the data that they are based upon. The benchmark calculations are also likely to show substantial dependence on choices of default parameters and default dose-response models. Ideally, these choices should be based on appropriate medical considerations, rather than on tradition. When limited information is available to guide the choice of conditions for the calculations, benchmark doses should therefore be calculated for a variety of relevant settings. Presentation of such results should stimulate discussions on the health implications associated with different model choices, and they should hopefully inspire further research in the field of dose-response curve modeling.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Pal Weihe for allowing us to use data from the Faroese cohort. This study was supported by grants from the National Institute of Environmental Health Sciences (ES06112 and ES09797), the U.S.Environmental Protection Agency (9W-0262-NAEX), the European Commission (Environment Research Programme), the Danish Medical Research Council, and the Danish Health Insurance Foundation. We thank Jørgen Hilden, the editors and referees for helpful comments.

REFERENCES

- Alho, J.M. and Valtonen, E. (1995). Interval estimation of inverse dose-response. *Biometrics* **51**, 491-501.
- Allen, B.C., Kavlock, R.J., Kimmel, C.A. and Faustman, E.M. (1994). Dose-response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with no observed effect levels. *Fundamental and Applied Toxicology* **23**, 487-495.
- Bailer, A.J., Stayner, L.T., Smith, R.J., Kuempel, E.D. and Prince, M.M. (1997). Estimating benchmark concentrations and other noncancer endpoints in epidemiology studies. *Risk Analysis* **17**, 771-780.
- Crump, K. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology* **4**, 854-871.

- Crump, K. (1995). Calculations of benchmark doses from continuous data. *Risk Analysis* **15**, 79-89.
- Crump, K., Kjellström, T., Shipp, A.M., Silvers, A. and Stewart, A. (1998). Influence of prenatal mercury exposure upon scholastic and psychological test performance: Benchmark analysis of a New Zealand cohort. *Risk Analysis* **18**, 701-713.
- Crump, K., Van Landingham, C., Shamlaye, C., Cox, C., Davidson, P.W., Myers, G.J. and Clarkson, T.W. (2000). Benchmark concentrations for methylmercury obtained from the Seychelles Child Development Study. *Environmental Health Perspectives* **108**, 257-263.
- Crump, K., Viren, J., Silvers, A., Clewel, H. III, Gearhart, J. and Shipp, A. (1995). Reanalysis of dose-response data from the Iraqi methylmercury poisoning episode. *Risk Analysis* **15**, 523-532.
- Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. London: Chapman & Hall.
- Environmental Protection Agency (EPA) (1997). *Mercury Study Report to Congress*. Washington, D.C.: U.S. Environmental Protection Agency.
- Grandjean, P. (1999). Mercury risks: Controversy or just uncertainty? *Public Health Reports* **114**, 512-515.
- Grandjean, P., Weihe, P., White, R.F., Debes, F., Araki, S., Yokoyama, K., Murata, K., Sørensen, N., Dahl, R. and Jørgensen, P.J. (1997). Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicology and Teratology* **19**, 417-428.
- International Programme on Chemical Safety (1994). *Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-Based Exposure Limits*. *Environmental Health Criteria* 170. Geneva: World Health Organization.
- Johnson, N.L. and Welch, B.L. (1939). Applications of the noncentral *t*-distribution. *Biometrika* **31**, 362-389.
- Kodell, R.L. and West, R.W. (1993). Upper confidence limits on excess risk for quantitative responses. *Risk Analysis* **13**, 177-182.
- National Academy of Sciences (NAS) (2000). *Toxicological Effects of Methylmercury*. Washington, D.C.: National Academy Press.
- Royston, P. and Altman, D.G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics* **43**, 429-467.