An extended and unified modeling framework for benchmark dose estimation for both continuous and binary data

Marc Aerts[1], **Matthew W. Wheeler**[2], **José Cortiñas Abrahantes**[3]

[1]Data Science Institute, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium

[2]National Institute for Occupational Safety and Health, Cincinnati, Ohio

[3]European Food Safety Authority, Parma, Italy

Abstract

Protection and safety authorities recommend the use of model averaging to determine the benchmark dose approach as a scientifically more advanced method compared with the no-observed-adverse-effect-level approach for obtaining a reference point and deriving health-based guidance values. Model averaging however highly depends on the set of candidate dose–response models and such a set should be rich enough to ensure that a well-fitting model is included. The currently applied set of candidate models for continuous endpoints is typically limited to two models, the exponential and Hill model, and differs completely from the richer set of candidate models currently used for binary endpoints. The objective of this article is to propose a general and wide framework of dose response models, which can be applied both to continuous and binary endpoints and covers the current models for both type of endpoints. In combination with the bootstrap, this framework offers a unified approach to benchmark dose estimation. The methodology is illustrated using two data sets, one with a continuous and another with a binary endpoint.

Keywords

Akaike information criterion; benchmark dose; bootstrap; cumulative distribution function; dose response; maximum likelihood; model averaging

1 | INTRODUCTION

It is widely agreed that the risk assessment of a specific substance at expected human exposure levels can be improved if extended use is made of the dose–response curves from studies in experimental animals for that substance, in order to better characterize

and quantify potential risks (Barlow et al., 2002; EFSA, 2009). Historically, when animal data have been used for risk assessment of chemicals in food, the no-observed-adverse-effect-level (NOAEL), the highest dose that does not differ significantly from the control, formed the reference point (RP) for the risk assessment, for deriving health-based guidance values, such as an acceptable daily intake (ADI). Disadvantages of this method are well known. Major drawbacks are that it is restricted to be one of the chosen experimental dose levels, and that it ignores any (monotone) shape of the dose–response curve. Crump (1984) developed a more quantitative approach, the benchmark dose (BMD) approach, based on dose response modeling techniques and the statistical lower confidence limit for the dose corresponding to a specific low but measurable target effect.

In its updated guidance (Hardy et al., 2017), the EFSA Scientific Committee confirms that the benchmark dose (BMD) approach is a scientifically more advanced method compared with the NOAEL approach for deriving a reference point. This document provides guidance on how to apply the BMD approach, and recommends model averaging as the preferred method for calculating the BMD confidence interval. The lower bound of this interval (BMDL) is needed as a potential RP, and the upper bound (BMDU) is needed for establishing the BMDU/BMDL ratio reflecting the uncertainty in the BMD estimate (Hardy et al., 2017). There is a vast literature on BMD estimation, and guidance and recommendations are provided by the European Food Safety Agency (EFSA) and the U.S. Environmental Protection Agency (EPA).

Here we focus on parametric dose–response models for continuous as well as binary data. Focusing on dose–response experiments with quantal data, West et al. (2012) studied, via large-scale Monte Carlo simulations, the effects of model selection and possible misspecification on the BMD and its corresponding lower confidence limit (BMDL). They called for further research in advanced statistical techniques incorporating the effects of model uncertainty, including model averaging approaches. Wheeler, Shao, and Bailer (2015) proposed an alternative reformulation of the BMD, based on milder assumptions using quantile regression and monotone smoothing splines. Haber et al. (2018) provides a review of the literature and the state-of-the-science with regard to BMD modeling. They also identify areas where consensus exists and areas where different perspectives exist, important in moving toward harmonization and a shared understanding of approaches. More literature on model averaging in the context of benchmark dose estimation has become available including Faes, Aerts, Geys, and Molenberghs (2007), Wheeler and Bailer (2007), Wheeler and Bailer (2008), Piegorsch et al. (2013), Shao and Gift (2014), and Simmons et al. (2015). User friendly software includes US EPA's BMDS software (Davis, Gift, & Zhao, 2011) and PROAST (see also EFSA, 2011).

Starting point of model averaging is determining the set of candidate models, which should contain a sufficiently large number of (non)nested models. The traditional set of candidate models, as included and described in EFSA's guidance, is (i) rather limited for continuous outcomes (exponential and Hill models) and (ii) quite different for continuous and binary endpoints (the latter also referred to as quantal data). The main objective of this article is to extend the family of candidate models considerably and to unify this family for both type of endpoints. A general framework of candidate models is defined, and it is shown that

the current models are included as special cases. Such unified framework is also expected to contribute to the harmonization of guidelines and recommendations by governmental authorities and agencies. Having a large set of (non)nested and sufficiently flexible candidate models can only improve the estimation of the benchmark dose based on model averaging. The use and performance of the new family of candidate models is illustrated on two data sets. The cell proliferation data (CP data, De Jong et al., 2002), with a continuous endpoint, are analyzed using the full design as well as a reduced design, adding some useful insights in the impact of the design on the BMD estimation. The thyroid epithelial cell vacuolization data (TECV data, Hardy et al., 2017) serve as an example for binary endpoints. For both illustrations we follow the estimation procedure of EFSA's guidance (see flowchart in figure 8 in Hardy et al., 2017; see also Figure 3 in the Supplementary Material).

The article is organized as follows. In Section 2 we introduce the new extended and unified modeling framework. General methodology on benchmark estimation is briefly described in Section 3. Section 4 summarizes the findings of a small simulation study. The CP- and TECV data are introduced and analyzed in Section 5, and a final Section 6 ends the article with discussion and topics for further research. Additional results and discussion are available in the Supplementary Material.

2 | A RICH UNIFYING FAMILY OF MODELS

Denote $y|x$ the endpoint of interest at a given dose x, having a Bernoulli distribution (for a binary endpoint) or a continuous distribution (typically the normal, possibly after a log transformation), with mean $E(y|x)$, denoted as $\pi(x)$ or as $\mu(x)$, respectively. For a binary endpoint one typically focuses on the "adverse" event such that it is very natural to assume that $\pi(x)$ is monotone increasing with dose x. For a continuous endpoint, it is very natural to assume it is monotone, but depending on the endpoint it might be monotone increasing or decreasing. For ease of notation and presentation and without loss of generality, we limit ourselves in this section mainly to a continuous and positive valued endpoint and to models for $\mu(x)$. But, everything applies to a binary endpoint with model $\pi(x)$ as well (see Section 2.7).

2.1 | General family for continuous endpoints

Consider a density function $f(x; b, \zeta)$ and corresponding cumulative distribution function $F(x; b, \zeta)$ of a positive random variable depending on a particular parameter b (playing the role of potency parameter, see next section) and possibly additional parameters ζ (suppressed from notation if not relevant). A new general family of dose–response models can be defined as, for $x > 0$

$$\mu_F(x) = a\{1 + (c - 1)F(x^d; b, \zeta)\},\tag{1}$$

where $a > 0$ without loss of generality, $b > 0$ possibly after reparameterization, $c > 1$ for an increasing dose–response model of a positive endpoint and likewise $c < 1$ for a decreasing dose–response model and finally $d > 0$. A reduced three-parameter model is obtained when fixing the fourth parameter $d = 1$. The case $d < 0$ is covered by considering the corresponding inverse distributions (see Section 2.6). The model can be further extended

with a fifth parameter ζ if available. There are two options, estimate ζ as any other parameter if the design and data allow to do so, or consider a series of models for a selection of particular fixed ζ values.

Note that the exponential model (Table 1) follows as a special case with $f(x;\ b) = be^{-bx}$, $F(x;\ b) = 1 - e^{-bx}$, being the exponential distribution with mean $1/b$, and the Hill model (Table 1) follows by choosing $f(x;\ b) = b/(b + x)^2, F(x;\ b) = x/(b + x)$ and reparameterizing b as b^d. In section 2.5.3 and figure 6 of Hardy et al. (2017), the graphical and practical interpretation of the four parameters a, b, c, and d, being the same for both the exponential and the Hill model, is illustrated and briefly discussed. In the next section, these interpretations are mathematically described and considered as desirable properties of the members of our general family.

2.2 | Properties parameterization existing models

The following properties are linking each of the parameters to specific characteristics of the dose–response curve. This eases interpretation of the dose–response model as well as the computation and interpretation of the BMD, in terms of these characteristics.

Property 1.—The parameter a equals the *background response*: $a = \mu(0)$.

Property 2.—The parameter c refers to the *maximum response*, relative to the background response: $c = \mu(\infty)/\mu(0)$. Note that $c > 1$ for an increasing dose response relationship and $c < 1$ in the opposite case. Furthermore, note that for a binary endpoint with $\pi(\infty) = 1$, this parameter is fixed as being $c = 1/a$ and can be left out.

Property 3.—The parameter b refers to the *potency*, determining that dose for which the steepness reaches its maximum; the solution x_M of $\frac{d^2\mu(x)}{d\log(x)^2}|_{x\ =\ x_M} = 0$ depends on b (and not on a or c).

Property 4.—The parameter d represents the *maximal steepness* (on log-dose scale): $\max_x \frac{d\mu(x)}{d\log(x)} \propto d$ with proportionality constant not depending on b.

Note that the currently used models, the exponential and Hill model (see Table 1), satisfy these properties. Indeed, straightforward calculations show that for the exponential model the above properties hold with $\max_x \frac{d\mu(x)}{d\log(x)} \propto d$ with proportionality constant $a(c - 1)/e$ (with e being Euler's number) not depending on b and with $x_M = b^{-1/d}$ not depending on a or c. For the Hill model the proportionality constant equals $a(c - 1)/4$ and the steepness reaches its maximum at $x_M = b^{1/d}$.

Properties 1 and 2 are trivially satisfied for the general family, and, as

$$\max_x \frac{d\mu_F(x)}{d\log(x)} = a(c - 1)d \max_u \left\{ f(u; b, \zeta)u \right\}.$$

Properties 3 and 4 hold if the choice of distribution F is constrained to a distribution for which the density f satisfies the condition that $\max_u\{f(u;b,\zeta)u\}$ does not depend on b, and with x_M being the solution of the differential equation

$$f'\left(x_M^d; b, \zeta\right)x_M^d + f\left(x_M^d; b, \zeta\right) = 0. \tag{2}$$

Different choices for the density function $f(x; b, \zeta)$ satisfying the condition that $\max_u\{f(u; b, \zeta)u\}$ does not depend on b, and with corresponding cumulative distribution function $F(x; b, \zeta)$, lead to different families of models. In the next section we propose three families, and their inverse versions as well: the gamma-based family extending the exponential model, the Lomax–Pareto-based family extending the Hill model, and the class of log-location-scale models.

2.3 | A gamma-based model family

The gamma distribution, for any $\zeta > 0$, has density

$$f(x; b, \zeta) = \frac{b^\zeta x^{\zeta - 1}e^{-bx}}{\Gamma(\zeta)}, \quad x > 0. \tag{3}$$

with cumulative distribution function

$$\Gamma(bx; \zeta) = \frac{\gamma(\zeta, bx)}{\Gamma(\zeta)},$$

with $\gamma(\zeta, u) = \int_0^u t^{\zeta - 1}e^{-t}dt$ the lower incomplete gamma function. This choice leads to the family of models

$$\mu_\Gamma(x) = a\left\{1 + (c - 1)\Gamma\left(bx^d; \zeta\right)\right\}, \quad x > 0. \tag{4}$$

Positive integer values for ζ correspond to the *Erlang distribution* with parameter ζ, as the sum of ζ independent exponentially distributed random variables with the same mean. For a general fixed (i.e., not estimated from the data) value $\zeta > 0$, the gamma distribution is also known as the *natural exponential* family.

Note that the choice $\zeta = 1$ leads to the exponential model. Some basic calculus shows that $x_M = (\zeta/b)^{1/d}$ and $\max_x \frac{d\mu_\Gamma(x)}{d\log(x)} = \left(a(c - 1)\frac{\zeta^\zeta}{\Gamma(\zeta)e^\zeta}\right)d$.

2.4 | A Lomax-based model family

As the gamma distribution generalizes the exponential distribution, the Lomax distribution arises as a mixture of exponential distributions, where the mixing distribution of the rate is a gamma distribution. The Lomax–Pareto distribution is a heavy-tail distribution (a shifted Pareto distribution) with density, for any $\zeta > 0$,

$$f(x; b, \zeta) = \frac{\zeta b^{\zeta}}{(x+b)^{\zeta+1}}, \quad x > 0, \tag{5}$$

resulting in the family of models

$$\mu_{LP}(x) = a\left\{1 + (c-1)\left[1 - \left(\frac{b}{b+x^d}\right)^{\zeta}\right]\right\}. \tag{6}$$

It can easily be shown that $x_M = (b/\zeta)^{1/d}$ and $\max_x \frac{d\mu_{LP}(x)}{d\log(x)} = \left(a(c-1)\left(\frac{\zeta}{1+\zeta}\right)^{\zeta+1}\right)d$. Note that the Lomax model reduces to the Hill model when $\zeta = 1$.

2.5 | A log-location-scale-based model

Consider the general family (1) with $F(x^d; b, \zeta) = F_{\zeta}(\ln(bx^d))$ for a specified cdf $F_{\zeta}(\cdot)$ and corresponding density $f_{\zeta}(\cdot)$, resulting in the dose–response model

$$\mu_{LLS}(x) = a\left\{1 + (c-1)F_{\zeta}(\ln b + d\ln x)\right\}, \quad x > 0. \tag{7}$$

The cdf $F_{\zeta}(\ln(bx^d))$ has the structure $F_{\zeta}(\frac{\ln(x)-\mu}{\sigma})$ with $\mu = -\ln b/d$ and $\sigma = 1/d$, known as the general form of a log-location-scale distribution (see Lawless, 2003) of a random variable X with $\ln(X)$ having mean μ, variance σ^2, and "standardized distribution" $F_{\zeta}(\cdot)$. Examples are the Weibull, log-logistic, and log-normal distribution. The Weibull and the log-logistic distribution do however not imply any new model $\mu(x)$, as the Weibull distribution results in the exponential model again, and the log-logistic in the Hill model. Here we focus on the log-normal distribution and its extension to the skewed-log-normal distribution, with one additional parameter ζ, being the skewness parameter (Azzalini, 1985), and leading to dose–response models denoted as $\mu_{LN}(x)$ and $\mu_{LSN}(x)$, respectively. The log-skew-normal with $\zeta = 0$ reduces to the log-normal model.

Some basic calculus shows that for the log-location-scale based model (7) $\max_u\{f(u; b, \zeta)u\} = \max_z\{f_{\zeta}(z)\}$ and the differential equation $f'(x_M^d; b, \zeta)x_M^d + f(x_M^d; b, \zeta) = 0$ translates to $f'_{\zeta}(\ln(bx_M^d)) = 0$ such that $x_M = (e^{M_{f\zeta}}/b)^{1/d}$ with $M_{f\zeta}$ the mode of the density $F_{\zeta}(\cdot)$ and $\max_x \frac{d\mu_{LLS}(x)}{d\log(x)} = (a(c-1)\max_z\{f_{\zeta}(z)\})d$.

2.6 | Families of inverse distributions

Consider a density function $f(x; b, \zeta)$ with corresponding cumulative distribution function $F(x; b, \zeta)$ of a positive random variable and the general family of dose response models $\mu_F(x) = a\{1 + (c-1)F(x^d; b, \zeta)\}$ as given in definition (1). The inverse distribution function of $F(x; b, \zeta)$ is given by the cdf $F_{IF}(x; b, \zeta) = 1 - F(x^{-1}; b, \zeta)$ and corresponding density by $f_{IF}(x; b, \zeta) = f(x^{-1}; b, \zeta)/x^2$, implying the following family of dose response models

$$\mu_{IF}(x) = a\left\{1 + (c - 1)\left(1 - F\left(x^{-d}; b, \zeta\right)\right)\right\}. \tag{8}$$

If the model family $\mu_F(x)$ satisfies Properties 1–4, also model family $\mu_{IF}(x)$ does. Indeed, it is easy to verify that $\max_u\{f_{IF}(u, b, \zeta)u\} = \max_u\{f(u, b, \zeta)u\}$ and that

$$f'_{IF}\left(x^d; b, \zeta\right)x^d + f_{IF}\left(x^d; b, \zeta\right) = -x^{-2d}\left\{f'\left(x^{-d}; b, \zeta\right)x^{-d} + f\left(x^{-d}; b, \zeta\right)\right\},$$

such that the zero point of the left-hand side equals the reciprocal x_M^{-1} of the zero point of Equation (2). Not each inverse distribution function leads to a new dose–response model. Indeed, the inverse of the Hill model and the log-normal model lead to the same respective model again.

2.7 | Further extensions and considerations

General family for binary endpoints—For a binary endpoint, the equivalent general family of dose–response models is obtained by fixing $c = 1/a$,

$$\pi_F(x) = a + (1 - a)F\left(x^d; b, \zeta\right). \tag{9}$$

The corresponding family based on the inverse distribution is also applicable in this setting.

Other families—Other distributions might be examined to satisfy Properties 1–4 and might be used as basis for a candidate response models, but we limit ourselves here to the models introduced above. It is expected that this set of candidate models is rich enough to contain at least one good fitting model to almost any dose–response data set. Distributions not satisfying Properties 1–4 may also be considered. Such a family is given by, assuming w.l.g. that $c > 0$ and $d > 0$,

$$\tilde{\mu}_F(x) = cF\left(a + bx^d\right), \quad x > 0, \tag{10}$$

for continuous endpoints and, fixing $c = 1$,

$$\tilde{\pi}_F(x) = F\left(a + bx^d\right), \quad x > 0, \tag{11}$$

for binary endpoints. Again the equivalent inverse distributions can be chosen as well. The interpretation of the parameters is less straightforward (e.g., mean response of the control group is $\tilde{\mu}_F(0) = cF(a)$ and depends on the model F) and the BMD typically depends on all parameters. In order to cover the existing logistic and probit model for binary endpoints, we will include those models in our full list of candidate models (for both type of endpoints).

Overview of the full family of models—Table 1 shows an overview of all candidate models considered further in the data analyses, their maximal number of parameters, their formula, and the expression for the BMD (see next section). This extensive family of models is fully available regardless the nature of the endpoint, continuous or binary. For the gamma distribution $\Gamma(\cdot; s, r)$, s refers to the shape and r to the rate parameter. As explained in the

next section, q^* refers to $\pm q/(c-1)$ for BMD definition (14) and to q for definitions (12) and (13). Note also that Table 1 covers all models that are recommended in the updated EFSA guidance (Hardy et al., 2017), albeit with Hill's model slightly differently parameterized (b for the Hill model in Table 1 to be reparameterized as b^d). The columns "Con" and "Bin" indicate whether the model is currently being used (y = yes), or not (n = not), for continuous and binary endpoints, respectively.

3 | BENCHMARK DOSE ESTIMATION

In this section we summarize key definitions in the field of BMD estimation and briefly describe model averaging within a maximum likelihood (ML) inferential framework. We opted for maximum likelihood inference, but the Bayesian paradigm could be chosen as well.

3.1 | Benchmark response, benchmark dose, and dose response models

The benchmark dose (BMD) is an active dose level associated with a specific relative increase (the benchmark response, BMR) compared with the mean response at zero dose. For a BMR of $q \times 100\%$ (e.g., $q = 0.01$), the BMD is defined as the solution of the equation

$$\frac{\pi(\text{BMD}) - \pi(0)}{\pi(\infty) - \pi(0)} = q, \tag{12}$$

for a binary response (in which case it is also natural to assume $\pi(\infty) = 1$), and likewise for a continuous endpoint as the solution of the equation

$$\frac{\mu(\text{BMD}) - \mu(0)}{\mu(\infty) - \mu(0)} = q. \tag{13}$$

For a continuous endpoint, the BMD is however often defined as the solution of the equation

$$\frac{\mu(\text{BMD}) - \mu(0)}{\mu(0)} = \pm q, \tag{14}$$

with $+q$ for an increasing and $-q$ for a decreasing dose–response relationship. Given data and a parametric model for $\pi(x)$ or $\mu(x)$, the model can be fitted and the BMD can be estimated from the data at hand. The BMDL and the BMDU are defined as the lower and upper (profile likelihood) confidence limits.

For the dose–response model (1), Equation (14) simplifies to

$$(c - 1)F\left(x^d; b, \zeta\right) = \pm q, \tag{15}$$

with solution

$$\text{BMD}_F = \left\{ F^{-1}\left(\frac{\pm q}{c-1}; b, \zeta\right) \right\}^{1/d}. \tag{16}$$

Note that the BMD_F in Equation (16) does depend on the choice of the cdf F, the BMR q, the model parameters b, c, d, and ζ, but not on parameter a. Equations (12) and (13) further simplify to

$$F\left(x^d; b, \zeta\right) = q, \tag{17}$$

with solution

$$\text{BMD}_F = \left\{ F^{-1}(q; b, \zeta) \right\}^{1/d}, \tag{18}$$

which, in contrast to the BMD_F defined by Equation (16), does not depend on c.

The formulas for the BMD for the family (10) or (11) have the general form:

$$\text{BMD}_F = \left\{ \frac{F^{-1}(\kappa) - a}{b} \right\}^{1/d}, \tag{19}$$

with

$$\kappa = \begin{cases} (1 \pm q)F(a) & \text{for definition (14),} \\ (1 - q)F(a) + q & \text{for definitions (12) and (13)}. \end{cases} \tag{20}$$

For the logistic model $F(u) = \text{expit}(u) = 1/(1 + e^{-u})$ and for the probit model $F(u) = \Phi(u)$ (the standard normal cumulative distribution function).

3.2 | Inference and model averaging

Consider a family $\mathscr{M} = \{\mu_k; k = 1, ..., K\}$ of candidate models. Consider a parameter of interest θ, common to all candidate models (e.g., the BMD). Fitting data $(x_1, y_1), ..., (x_n, y_n)$ to each candidate model μ_k provides a model specific ML-estimate $\hat{\theta}_k$, with standard-error estimate $\widehat{\text{se}}(\hat{\theta}_k)$. Instead of focusing on the estimate of the single best model using a model selection criterion such as Akaike's information criterion (AIC, Akaike, 1973) and thereby ignoring the data-driven model selection process in any further inference, model averaging accounts for that process and provides "multimodel inference." The model-averaged estimate is defined as

$$\hat{\theta}_{\text{MA}} = \sum_{k=1}^{K} w_k \hat{\theta}_k, \tag{21}$$

with weights w_k

$$w_k = \frac{\exp(-\Delta_k/2)}{\sum_{k=1}^{K} \exp(-\Delta_k/2)}, \tag{22}$$

where $\quad_k = \text{AIC}_k - \text{AIC}_{\min}$, the (corrected) AIC value of model k minus the minimal AIC over all candidate models in the set $\mathscr{M}$. An estimator for the standard error of $\hat{\theta}_{\text{MA}}$ is given by

$$\widehat{\text{se}}(\hat{\theta}_{\text{MA}}) = \sum_{k=1}^{K} w_k \sqrt{\widehat{\text{se}}^2(\hat{\theta}_k) + \left(\hat{\theta}_k - \hat{\theta}_{\text{MA}}\right)^2}, \tag{23}$$

combining within and between model variability, in this way accounting for model uncertainty as well as sampling variability. For more information on model averaging, see Anderson and Burnham (2002).

For estimating the BMD and determining the BMDL and BMDU, several approaches can be taken. In a first "direct" approach, the model-specific BMD estimates $\hat{\theta}_k = \widehat{\text{BMD}}_k$ for the common parameter $\theta = \text{BMD}$ are averaged directly, resulting in

$$\widehat{\text{BMD}}_{\text{DMA}} = \sum_{k=1}^{K} w_k \widehat{\text{BMD}}_k, \tag{24}$$

and leading to the BMDL and BMDU bounds

$$\widehat{\text{BMD}}_{\text{DMA}} \pm z_{\alpha/2} \widehat{\text{se}}(\widehat{\text{BMD}}_{\text{DMA}}), \tag{25}$$

with $z_{\alpha/2}$ the $(1 - \alpha/2)$-quantile of the standard normal distribution and with $\widehat{\text{se}}(\widehat{\text{BMD}}_{\text{DMA}})$ computed according to formula (23).

A Wald type of confidence interval might however lead to a negative value for the BMDL. Moreover, concerns have been raised about the accuracy of the standard error estimate (23) and bootstrap approaches have been proposed to overcome these shortcomings (see Faes et al., 2007). A bootstrap approach would require the generation of bootstrap samples. As the design is typically fixed (as in our data examples in Section 5), resampling the pairs $(x_1, y_1), \ldots, (x_n, y_n)$ is not appropriate. Here we illustrate the use of the parametric bootstrap, but semiparametric residual bootstrap is an option as well. For each observed dose level x_i, generate a bootstrap response value y_i^* as

$$y_i^* \mid x_i \sim f_{y\mid x}(y; \hat{\mu}(x_i), \hat{\sigma}), \quad i = 1, \ldots, n, \tag{26}$$

with $f_{y\mid x}(y; \mu, \sigma)$ being an appropriate density with parameters mean μ and variance σ^2 (often the normal distribution, possibly after a log transformation, for continuous endpoints; the binomial distribution for binary endpoints), with $\hat{\mu}(x_i)$ being the estimate based on the observed average response at dose x_i (i.e., ANOVA-estimate) or alternatively the model averaged response dose relationship $\hat{\mu}_{MA}(x_i)$ with

$$\hat{\mu}_{\text{MA}}(x) = \sum_{k=1}^{K} w_k \hat{\mu}_k(x). \tag{27}$$

The estimate $\hat{\sigma}^2$ can similarly be based on the ANOVA-residuals or alternatively on the residuals $y_i - \hat{\mu}_{MA}(x_i), i = 1, \ldots, n$. The choice using $\hat{\mu}_{MA}(x_i)$ and corresponding residuals is illustrated in the data examples (Section 5). For each $b = 1, \ldots, B$ of the B bootstrap samples, the "direct" BMD estimates $\widehat{\text{BMD}}_{DMA}^{(b)}$ are computed, thus generating the empirical bootstrap distribution from which the percentile bootstrap confidence limits can be derived (or alternatives thereof, see Davison and Hinkley 2006). In our data examples $B = 3,000$. In summary, the first "direct" approach leads to the point estimate (24) with interval bounds BMDL and BMDU determined based on asymptotic normality (Wald type) or alternatively based on bootstrap percentiles.

A second "indirect" approach goes as follows. Instead of averaging the model specific BMD estimates, one determines the single BMD corresponding to the averaged dose–response model $\hat{\mu}_{MA}(x)$ (defined by (27)). So, determine the $\widehat{\text{BMD}}_{IMA}$ as the solution of the equation (or similar expressions when using the versions (12) or (13))

$$\frac{\hat{\mu}_{MA}(\widehat{\text{BMD}}_{IMA}) - \hat{\mu}_{MA}(0)}{\hat{\mu}_{MA}(0)} = \pm q. \tag{28}$$

As a Wald-type confidence interval is not an attractive option for this approach, we only consider the bootstrap-based percentile intervals for this indirect method. This latter approach was also taken by Wheeler and Bailer (2007), who examined the properties of a model-averaged BMDL for binary endpoints.

4 | SIMULATION STUDY

For continuous outcomes, the current set of models is typically limited to the exponential and Hill models (Hardy et al., 2017). In a small simulation study it was examined how model averaging over the full family of candidate models (all models in Table 1) performs compared with the reduced set (exponential and Hill models only). The dose levels were set on $d = 2^{-\ell}$ for $\ell \in \{0, 1, 2, 3, 4, \infty\}$. At zero dose 30 response values and at each active dose level 15 response values were generated according to the normal distribution with mean $\mu(d)$ and variance $\sigma^2 = 1$. Figure 1 shows four different spline based dose response models for the mean $\mu(d)$ (Models A–D) with one generated sample (R code generating the data for the different models is available upon request from the authors). For each model, 2,000 samples were generated. Four other dose response models were investigated, but results were very similar to those of models A and B, and therefore results are not shown. Models D and C are related as $\mu_D(d) = \mu_C(d + 0.1)$, implying a slow (rapid) increase in response for $d \in [0, 0.1]$ for model C (model D).

Each of the candidate models were fit to the data, and the BMD, with $q = 0.01$ in Equation (14), was estimated based on model averaging, only using the reduced set and using the full set of candidate models. For nested models, only the model with lowest AIC was included. Across all simulations, the indirect method for estimating the BMD outperformed consistently (but with varying degrees) the direct approach. Therefore, we focus on the

comparison between both families of candidate models (reduced vs. full) for the indirect method only.

The R statistical computing environment (R Core Team, 2017) was used for implementing and fitting the models, and the R package AICcmodavg for computing model-averaged parameter estimates. Table 2 shows the performance in terms of mean squared error for each of the four dose response models A–D. Using the full family generally reduces the bias. For the variance it varies, but often it is a bit smaller or comparable, and in some other cases somewhat larger. The combination in terms of mean squared error (mse) is in general in favor of the full family. Interestingly, among all models we tried, only in case of model D, averaging over the full family performs less good. However, for that model D, the mse is very low for both sets of candidate models and a closer inspection shows that the distribution of the BMD estimates is very skewed to the right in this case, much more extreme than for all other scenarios, hampering the use of mse as a measure of accuracy. Averaging over the reduced family leads to the following characteristics of the BMD estimates of the BMD = 0.00146: 0.00135 (mean), 0.00092 (median), and 0.00011 (5% quantile, being the BMDL); averaging over the full family: 0.00299 (mean), 0.00146 (median), and 0.00011 (5% quantile, being the BMDL). So, the 5% quantiles are the same, and the median of the full family approach is very close to the true value.

5 | DATA EXAMPLES

Two data sets will be used to illustrate the proposed methodology, the cell proliferation (CP) data with a continuous endpoint, and the thyroid epithelial cell vacuolization data (TECV) data with a binary endpoint. The CP and TECV data are fitted with all candidate models (as listed in Table 1). The EFSA flowchart (see figure 8 in (Hardy et al., 2017), see also Figure 3 in the Supplementary Material) has been followed. Of course, there is no unique best way to guide the user through such a modeling and estimation exercise, but, in our view. this chart reflects statistical considerations as well as practical ones (Hardy et al., 2017). Of course, here we follow, at the bottom of the chart, the flow to the left as software is available (also upon request from the authors). For both data sets, the model-specific BMD estimates and 90% profile likelihood-based BMDL and BMDU limits are calculated, as well as the model-averaged BMD estimates based on the direct approach with 90% Wald and percentile bootstrap BMDL and BMDU limits, and based on the indirect approach with 90% percentile bootstrap BMDL and BMDU limits. Also estimates for the standard error are shown: using formula (23) as well as using the bootstrap for the direct estimate $\widehat{\mathrm{BMD}}_{\mathrm{DMA}}$ and only the bootstrap for the indirect estimate $\widehat{\mathrm{BMD}}_{\mathrm{IMA}}$. All bootstrap results are based on $B = 3,000$ bootstrap samples.

5.1 | Cell proliferation data

De Jong et al. (2002) applied the local lymph node assay (LLNA, OECD guideline 406) test to 15 different rubber chemicals. Here we focus on one of the rubber chemicals. Test chemicals are applied to the dorsum of the ear and lymphocyte activation is determined by measuring cell proliferation in the draining auricular lymph nodes. For more details on the design of the study, see De Jong et al. (2002). Slob and Setzer (2014) reanalyzed these data

indicating that an exponential or a Hill model with four parameters adequately describes toxicological dose responses. We will analyze the log-transformed values.

The CP data are analyzed with the full design (all data, left panel of Figure 2) as well as with a reduced design (Figure 4 in the Supplementary Material). Analyzing the data according to both designs provides some further insights of the impact of the design of the study, on the extent to which the fits of all candidate models are affected by the design and on whether the model-averaged estimate is robust or highly sensitive to the design (see Supplementary Material). Here we focus on the full design. The left panel of Figure 2 shows the individual fitted models and corresponding BMD estimates (red vertical lines on the x-axis) to the CP data. Table 3 shows the model-specific estimates, the (corrected) AIC values, and corresponding weights; the model-averaged BMD estimates and the se-estimates based on formula (23) as well as on the bootstrap, and finally Wald-type confidence limits as well as percentile bootstrap limits.

The "full" ANOVA model has an AIC value $\text{AIC}_{\text{FULL}} = 49.319$, and for the constant mean "null" model $\text{AIC}_{\text{NULL}} = 116.012$. As Table 3 shows the model with lowest AIC has an AIC $= 50.647 \quad \text{AIC}_{\text{NULL}}$-2 (actually all models have), the data are considered suitable for calculating a BMD and a BMDL/U (following the EFSA flowchart). Testing the residuals of the ANOVA model with the Shapiro–Wilk normality test indicates no evidence against the normality assumption (p-value = .417). Estimates are shown in Table 3. The estimates of the gamma model are not included, as this model did not fit better than the exponential model (according to AIC). Likewise, the results of the inverse gamma, Lomax, inverse Lomax, log-skew-normal, and inverse log-skew-normal are not included, as the corresponding simpler nested model showed a lower AIC. The logistic and probit model resulted in the lowest AIC, receiving consequently the highest weights, but closely followed by the exponential model. These three models get weights in the range 0.20–0.22. The other three remaining models receive weights in the approximate range 0.1–0.15. The model-specific BMD estimates vary from 0.07 to 0.13, the BMDL values from 0.01 to 0.09, and the BMDU values from 0.31 to 0.45. The direct approach results in the averaged BMD estimate $\widehat{\text{BMD}}_{\text{DMA}} = 0.091$ with Wald-based limits BMDL = 0.014 and BMDU = 0.409 and bootstrap-based limits BMDL = 0.011 and BMDU = 0.325. The indirect approach results in the same averaged BMD estimate $\widehat{\text{BMD}}_{\text{IMA}} = 0.091$ with bootstrap-based limits BMDL = 0.019 and BMDU = 0.318.

5.2 | Thyroid epithelial cell vacuolization data

This example relates to a 2-year study in rats, where three active doses of a compound were administered to the animals. Dose-related changes in thyroid epithelial cell vacuolization (TECV) were found. The TECV data are shown in Table 4 and corresponding proportions are graphically depicted as a function of dose (on log(dose+1) scale) in the right panel of Figure 2. These data were used to illustrate the recommended procedure in the earlier EFSA guidance (EFSA, 2009) as well as in its recent 2017 update (Hardy et al., 2017).

For the TECV data we have $\text{AIC}_{\text{FULL}} = 188.039$ and $\text{AIC}_{\text{NULL}} = 276.372$, so that again all models have an AIC $\quad \text{AIC}_{\text{NULL}}$-2 (see Table 5). For this binary endpoint, the inverse exponential model shows the lowest AIC, followed by the Lomax and the log-skew-normal

model. The logistic and probit models are fitting less well compared with the other candidate models, and consequently get very low weight in the model average procedure. The direct approach results in the averaged BMD estimate $\widehat{\text{BMD}}_{\text{DMA}} = 2.443$ with Wald-based limits BMDL = 0.295 and BMDU = 4.590 and bootstrap based limits BMDL = 0.707 and BMDU = 3.734. The indirect approach results in the averaged BMD estimate $\widehat{\text{BMD}}_{\text{IMA}} = 2.153$ with bootstrap based limits BMDL = 0.261 and BMDU = 3.074.

6 | DISCUSSION AND FURTHER RESEARCH

The proposed suite of parametric candidate models extends and unites the existing distinct ones for continuous and binary endpoints. Irrespective of the endpoint, the same family of dose response models can be applied as individual models on their own or as models contributing to the averaged dose response model and averaged BMD estimate. The suite is defined in a generic way with parameters sharing common response characteristics and properties. A particular set of members, based on the cumulative distribution function of the gamma, the Lomax and log-normal distribution is shown to satisfy these properties and to cover most of the models typically used in daily practice. Whereas focus here was on individual data and the use of the (log)normal distribution describing the variation of these data about the mean dose response, the models can be applied on aggregated data and in combination with other distributions as well.

Two approaches for model-averaged BMD estimation are considered: the "direct" approach averaging the model specific BMD estimates and the "indirect" approach, first averaging the dose response models and next using the averaged model to obtain the final BMD estimate. A small simulation study for a continuous endpoint confirms that the use of the proposed extended suite of models improves the estimation of the BMD, in line with what is generally expected and known from literature on model averaging.

There are several avenues for further research. First of all, the use of other cumulative distributions could be explored. Based on the findings of Wheeler and Bailer (2007), the use of the indirect approach can be recommended for binary endpoints, but more research is needed to reveal additional insights in the differences in performance between the direct and indirect model averaging approach across different endpoints and different settings. Extensive simulations are needed to investigate the performance of the proposed suite of models for a variety of studies, with different sample sizes, designs, endpoints, and so forth. For sparse designs with only a few distinct dose levels, parametric dose response models with more than three parameters may be overparameterized and may cause nonconvergence and related computational issues. Penalized likelihood might offer an interesting solution to this particular issue. Currently, we are also looking into extending the methodology to include covariates and to deal with hierarchical and clustered data structures (such as endpoints from fetuses clustered within litters).

Wald-type confidence intervals are known to be less optimal (in terms of their reliance on the asymptotic normal distribution, their coverage, and their compliance with parameter constraints). Bootstrap percentile confidence intervals are expected to circumvent most of these shortcomings but are very computer intensive. Fletcher and Turek (2011) proposed a

method for model averaging a set of single-model profile likelihood intervals, making use of the link between profile likelihood intervals and Bayesian credible intervals. This approach could be explored further.

Currently we are implementing a Bayesian version, providing Bayesian credible intervals and a natural way to incorporate other information through the prior distribution of model parameters, which is expected to be beneficial when performing dose–response modeling for noninformative or minimally informative data (which may be the only data available for risk assessors). Fang, Piegorsch, and Barnes (2015) describe a hierarchical Bayesian approach for estimating BMDs in quantitative risk analysis, with emphasis on environmental carcinogenicity assessment. They also mention that model adequacy is an important issue in benchmark risk analysis. Model averaging offers a solution to model adequacy and uncertainty and multimodel inference has a long history within the Bayesian paradigm. Some recent developments in Bayesian model averaging offer promising ways of implementation (see Yao, Vehtari, Simpson, & Gelman, 2017). We hope to report on Bayesian model-averaged BMD estimation in a future manuscript.

Whitney and Ryan (2013) proposed a new method for BMD risk estimation adapted to epidemiological data (with typically all observed dose levels being positive), by specifying a nonzero background exposure level. To account for uncertainty in that level, their modified BMDs are averaged over a distribution of exposure levels. It would be very interesting to examine how our unified modeling framework can be applied to this epidemiological setting as well.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The cell proliferation data that illustrate the proposed methodology and findings of this study for continuous endpoints are expected to be available from De Jong et al. (2002) upon reasonable request.

REFERENCES

Akaike H (1973). Information theory and an extension of the maximum likelihood principle (pp. 199–213). New York, NY: Springer.

Anderson DR, & Burnham KP (2002). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). New York, NY: Springer.

Azzalini A (1985). A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12, 171–178.

Barlow S, Dybing E, Edler L, Eisenbrand G, Kroes R, & Brandt PA (2002). Food safety in Europe (FOSIE): Risk assessment of chemicals in food and diet. Food and Chemical Toxicology, 01, 40.

Crump KS (1984). A New Method for Determining Allowable Daily Intakes. Fundamental and Applied Toxicology, 4, 854–871. [PubMed: 6510615]

Davis JA, Gift JS, & Zhao QJ (2011). Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1. Toxicology and Applied Pharmacology, 254(2), 181–191. [PubMed: 21034758]

Davison AC, & Hinkley DV (2006). Bootstrap methods and their application Cambridge series in statistical and probabilistic mathematics. Cambridge, MA: Cambridge University Press.

De Jong WH, Van Och FMM, Den Hartog Jager CF, Spiekstra SW, Slob W, Vandebriel RJ, & Van Loveren H (2002). Ranking of allergenic potency of rubber chemicals in a modified local lymph node assay. Toxicological Sciences, 66(2), 226–232. [PubMed: 11896289]

EFSA. (2009). Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. EFSA Journal, 1150, 1–72.

EFSA. (2011). Use of BMDS and PROAST software packages by EFSA scientific panels and units for applying the benchmark dose (BMD) approach in risk assessment. EFSA Supporting Publications, 8(2), 113E.

Faes C, Aerts M, Geys H, & Molenberghs G (2007). Model averaging using fractional polynomials to estimate a safe level of exposure. Risk Analysis, 27(1), 111–123. [PubMed: 17362404]

Fang Q, Piegorsch WW, & Barnes KY (2015). Bayesian benchmark dose analysis. Environmetrics, 26(5), 373–382. 10.1002/env.2339

Fletcher D, & Turek D (2011). Model-averaged profile likelihood intervals. Journal of Agricultural, Biological, and Environmental Statistics, 17(1), 38–51. 10.1007/s13253-011-0064-8

Haber LT, Dourson ML, Allen BC, Hertzberg RC, Parker A, Vincent MJ, … Boobis AR (2018). Benchmark dose (BMD) modeling: current practice, issues, and challenges. Critical Reviews in Toxicology, 48(5), 387–415. [PubMed: 29516780]

Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen KH, More S, … Ockleford C (2017). Update: Use of the benchmark dose approach in risk assessment. EFSA Journal, 15(1), e04658. 10.2903/j.efsa.2017.4658 [PubMed: 32625254]

Lawless JF (2003) Statistical models and methods for lifetime data. New Jersey, USA: John Wiley and Sons, 2nd edn.

Piegorsch WW, An L, Wickens AA, Webster West R, Pena EA, & Wu W (2013). Information-theoretic model-averaged benchmark dose analysis in environmental risk assessment. Environmetrics, 24(3).143–157. doi:10.1002/env.2201. [PubMed: 24039461]

R Core Team. (2017. Retrieved from). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org

Shao K, & Gift JS (2014). Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. Risk Analysis, 34(1), 101–120. [PubMed: 23758102]

Simmons S, Chen C, Li X, Wang Y, Piegorsch W, Fang Q, … Dunn GE (2015). Bayesian model averaging for benchmark dose estimation. Environmental and Ecological Statistics, 22(1), 5–16.

Slob W, & Setzer RW (2014). Shape and steepness of toxicological doseresponse relationships of continuous endpoints. Critical Reviews in Toxicology, 44(3), 270–297. [PubMed: 24252121]

West RW, Piegorsch WW, Peña EA, An L, Wu W, Wickens AA, … Chen W (2012). The impact of model uncertainty on benchmark dose estimation. Environmetrics, 23(8), 706–716. 10.1002/env.2180 [PubMed: 23794799]

Wheeler MW, & Bailer AJ (2007). Properties of model-averaged BMDLs: A study of model averaging in dichotomous response risk estimation. Risk Analysis, 27(3), 659–670. 10.1111/j.1539-6924.2007.00920.x [PubMed: 17640214]

Wheeler MW, & Bailer AJ (2008). Model averaging software for dichotomous dose response risk estimation. Journal of Statistical Software, 26(5), 1–15. [PubMed: 19777145]

Wheeler MW, Shao K, & Bailer AJ (2015). Quantile benchmark dose estimation for continuous endpoints. Environmetrics, 26(5), 363–372. 10.1002/env.2342

Whitney M, & Ryan L (2013). Uncertainty due to low-dose extrapolation: Modified BMD methodology for epidemiological data. Environmetrics, 24(5), 289–297. 10.1002/env.2217

Yao Y, Vehtari A, Simpson D, & Gelman A (2017). Using stacking to average Bayesian predictive distributions. Bayesian Analysis, 04, 13.
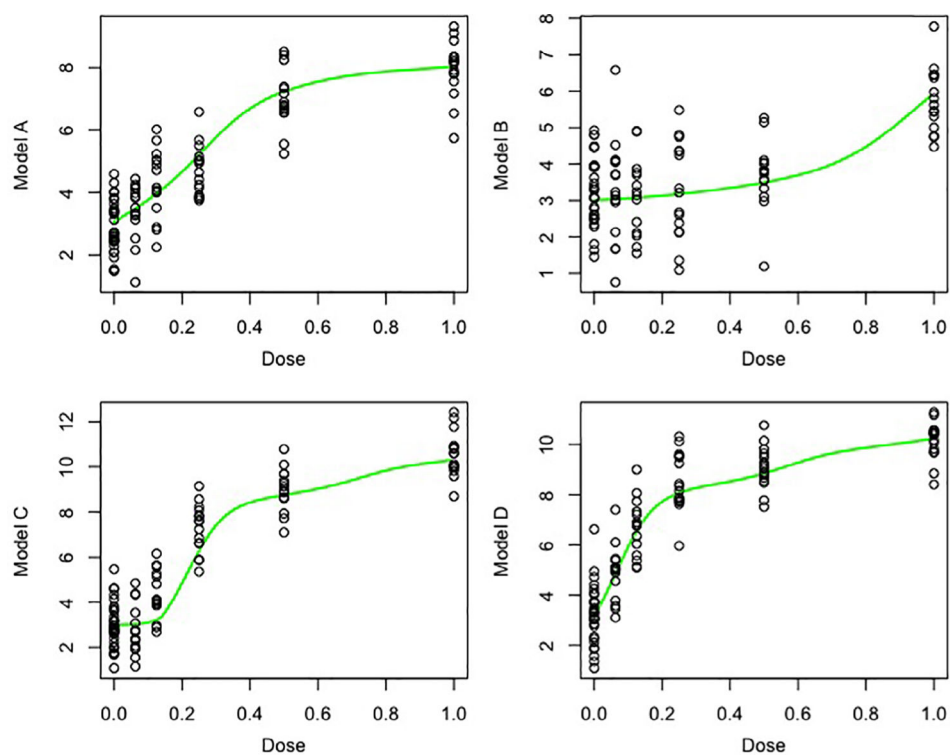
FIGURE 1.
The dose response models of the simulations study overlaid with one generated data set of a normally distributed continuous endpoint.
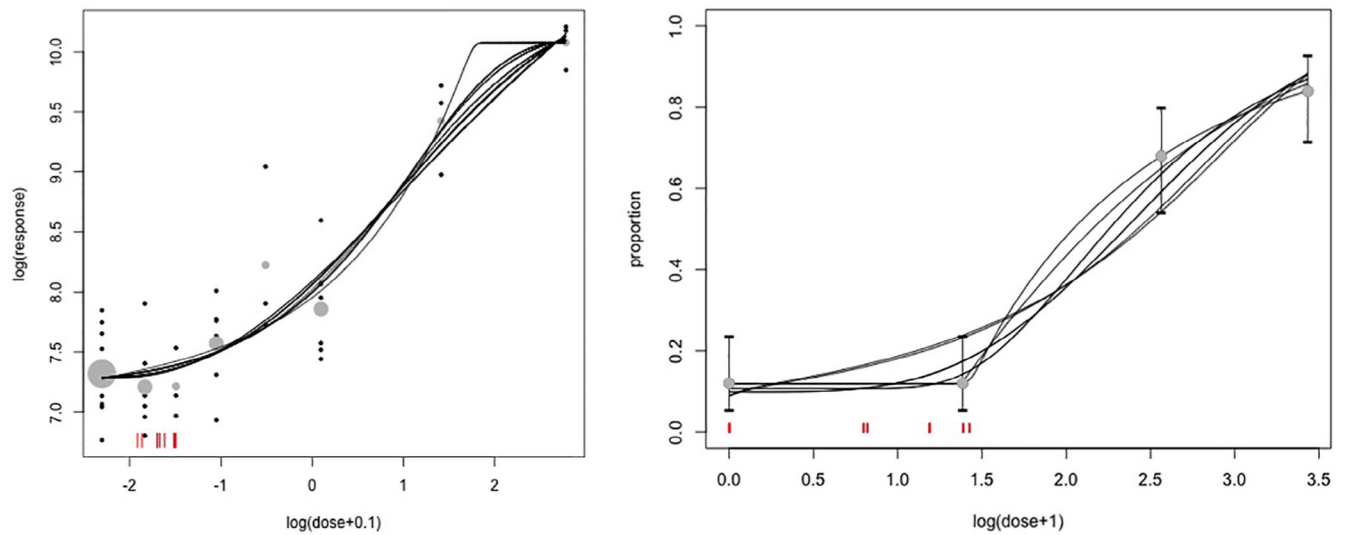
FIGURE 2.
Left: the cell proliferation data with full design, model-specific dose response models, and estimated BMDs (short red vertical lines along and above the horizontal axis). Right: the thyroid epithelial cell vacuolization data, model specific dose response models, and estimated BMD's (short red vertical lines along and above the horizontal axis)

TABLE 1

Table of candidate models, with columns: the name of the model; the number of parameters "# pars"; an indicator (y = yes and n = no) of its current use in the EFSA guidelines (Hardy et al., 2017), in the column "Con" for continuous and the column "Bin" for binary endpoints; the constraint $c = 1/a$ or $c = 1$ for binary endpoints (column "c for Bin"); the mathematical formula for the model; and the mathematical formula for the BMD

Name	# pars	Con[b]	Bin[b]	c for Bin	Dose-response model	BMD[c]
Exponential	4	y	n	$\frac{1}{a}$	$a\{1 + (c-1)(1 - \exp(-bx^d))\}$	$\left\{\frac{-\ln(1 - q^*)}{b}\right\}^{1/d}$
Inverse exponential	4	n	y	$\frac{1}{a}$	$a\{1 + (c-1)\exp(-bx^{-d})\}$	$\left\{\frac{-b}{\ln(q^*)}\right\}^{1/d}$
Gamma	5	n	y	$\frac{1}{a}$	$a\{1 + (c-1)\Gamma(bx^d;\zeta)\}$	$\left\{\frac{\Gamma^{-1}(q^*;\zeta)}{b}\right\}^{1/d}$
Inverse gamma	5	n	n	$\frac{1}{a}$	$a\{1 + (c-1)(1 - \Gamma(bx^{-d};\zeta))\}$	$\left\{\frac{b}{\Gamma^{-1}(1 - q^*;\zeta)}\right\}^{1/d}$
Hill $\overset{a}{=}$ Inverse Hill	4	y	y	$\frac{1}{a}$	$a\{1 + (c-1)(1 - \frac{b}{b + x^d})\}$	$\left\{\left(\frac{1}{1 - q^*} - 1\right)b\right\}^{1/d}$
Lomax	5	n	n	$\frac{1}{a}$	$a\left\{1 + (c-1)(1 - \left(\frac{b}{b + x^d}\right)^\zeta)\right\}$	$\left\{\left(\frac{1}{(1 - q^*)^{1/\zeta}} - 1\right)b\right\}^{1/d}$
Inverse Lomax	5	n	n	$\frac{1}{a}$	$a\{1 + (c-1)\left(\frac{b}{b + x^{-d}}\right)^\zeta\}$	$\left\{\left(\frac{1}{((q^*)^{-1/\zeta} - 1)b}\right)\right\}^{1/d}$
Log-normal $\overset{a}{=}$ Inverse log-normal	4	n	y	$\frac{1}{a}$	$a\{1 + c - 1)\Phi(\ln b + d\ln x)\}$	$\exp\left\{\frac{\Phi^{-1}(q^*) - \ln b}{d}\right\}$
Log-skew-normal	5	n	n	$\frac{1}{a}$	$a\{1 + c - 1)\Phi_{SN}(\ln b + d\ln x; \zeta)\}$	$\exp\left\{\frac{\Phi_{SN}^{-1}(q^*;\zeta) - \ln b}{d}\right\}$
Inverse log-skew-normal	5	n	n	$\frac{1}{a}$	$a\{1 + c - 1)(1 - \Phi_{SN}(\ln b - d\ln x; \zeta))\}$	$\exp\left\{\frac{\ln b - \Phi_{SN}^{-1}(1 - q^*;\zeta)}{d}\right\}$
Logistic	4	n	y	1	$c\,\mathrm{expit}(a + bx^d)$	$\left\{\frac{\mathrm{logit}(\kappa) - a}{b}\right\}^{1/d}$
Probit	4	n	y	1	$c\Phi(a + bx^d)$	$\left\{\frac{\Phi^{-1}(\kappa) - a}{b}\right\}^{1/d}$

[a] The inverse of the Hill model and the log-normal model lead to the same respective model again.

[b] y = yes (currently used) and n = no (currently not used).

[c] q^* refers to $\pm q/(c - 1)$ for BMD definition (14) and to q for definition (12) and (13) and κ is defined by (20).

TABLE 2

Simulation study with Models A–D, the value of the true BMD, squared bias (×100) when using only the exponential and Hill models (R = reduced set) and when using the full family (F), likewise the variance (×100) and mean squared error (×100) for both sets, with a indicating the lowest values

Model	BMD	Bias²-R	Bias²-F	Var-R	Var-F	mse-R	mse-F
A	0.0048	0.0260	0.0141[a]	0.0182	0.0172[a]	0.0442	0.0314[a]
B	0.0604	4.4878	4.4095[a]	3.0596	2.5476[a]	7.5474	6.9571[a]
C	0.0709	0.1216	0.0014[a]	0.0167[a]	0.0606	0.1383	0.0621[a]
D	0.0015	0.0000[a]	0.0002	0.0002[a]	0.0014	0.0002[a]	0.0016

TABLE 3

The cell proliferation data with full design

Model	a	b	c	d	ζ	σ	BMD	BMDL	BMDU	AIC	Weight
Exp	7.285 (0.097)	0.322 (0.085)	1.388 (0.040)	1.012 (0.228)	—	0.394 (0.043)	0.083 (0.065)	0.025	0.333	50.859	0.197
Inv Exp	7.284 (0.097)	1.696 (0.478)	1.654 (0.289)	0.434 (0.188)	—	0.400 (0.044)	0.125 (0.085)	0.091	0.449	52.207	0.101
Hill	7.288 (0.098)	2.937 (0.931)	1.435 (0.075)	1.1524 (0.334)	—	0.396 (0.043)	0.098 (0.081)	0.014	0.401	51.447	0.147
LN	7.289 (0.097)	0.525 (0.111)	1.445 (0.099)	0.646 (0.204)	—	0.397 (0.043)	0.122 (0.087)	0.017	0.439	51.738	0.127
Logistic	0.949 (0.100)	0.420 (0.107)	10.103 (0.237)	0.956 (0.214)	—	0.393 (0.043)	0.077 (0.061)	0.011	0.320	50.758	0.208
Probit	0.588 (0.057)	0.245 (0.061)	10.089 (0.226)	0.920 (0.197)	—	0.393 (0.043)	0.071 (0.057)	0.011	0.306	50.647	0.220
The direct approach: standard errors estimate (23) and Wald-type limits							0.091 (0.073)	0.014	0.409		
The direct approach: bootstrap standard error and percentile limits							0.091 (0.103)	0.011	0.325		
The indirect approach: bootstrap standard error and percentile limits							0.091 (0.101)	0.019	0.318		

Note: Estimates for all fitted models (with one representative for nested models), the model-specific BMD estimates, 90% profile likelihood based BMDL and BMDU, corrected AIC, and corresponding weight for averaging. Model-averaged BMD estimates based on the direct approach with 90% Wald and percentile bootstrap BMDL and BMDU limits, and based on the indirect approach with 90% percentile bootstrap BMDL and BMDU limits. Estimates for the standard error are shown: using formula (23) and the bootstrap for the direct estimate and only the bootstrap for the indirect estimate. $B = 3{,}000$ bootstrap samples were generated.

Abbreviations: Exp, exponential; Inv Exp, inverse exponential; LN, log-normal; LSN, log-skew-normal model.

TABLE 4

The thyroid epithelial cell vacuolization data: dose level in milligram per kilogram birth weight, per day; number of animals showing thyroid epithelial cell vacuolization; number of animals in each dose group

Dose (mg kg bw^{-1} day^{-1})	# Animals showing TECV	# Animals in dose group
0	6	50
3	6	50
12	34	50
30	42	50

TABLE 5

The thyroid epithelial cell vacuolization data

	a	b	d	ζ	BMD	BMDL	BMDU	AICc	Weight
Inv Exp	0.107 (0.037)	13.628 (8.707)	1.312 (0.246)	—	2.287 (0.785)	1.298	4.127	187.262	0.347
Gamma	0.096 (0.035)	6.377 (5.511)	0.297 (0.095)	13.392 (8.645)	1.225 (0.536)	0.525	2.340	192.065	0.031
Lomax	0.100 (0.034)	108.112 (126.342)	3.028 (1.253)	0.324 (0.232)	3.164 (2.062)	0.940	7.024	188.244	0.212
Inverse Lomax	0.106 (0.037)	1.925 (7.024)	1.328 (0.255)	27.721 (94.074)	2.282 (0.784)	1.313	4.126	189.351	0.122
LN	0.099 (0.036)	0.075 (0.038)	1.090 (0.184)	—	1.271 (0.547)	0.562	2.434	189.850	0.095
LSN	0.117 (0.041)	0.517 (0.098)	0.606 (0.084)	40.200 (179.513)	3.009 (0.487)	1.989	3.988	188.591	0.179
Logistic	−2.327 (0.434)	0.594 (0.292)	0.586 (0.132)	—	0.001 (0.003)	0.001	0.029	194.904	0.008
Probit	−1.337 (0.218)	0.322 (0.151)	0.604 (0.128)	—	0.001 (0.003)	0.001	0.029	195.719	0.005
The direct approach: standard errors estimate (23) and Wald-type limits					2.443 (1.306)	0.295	4.590		
The direct approach: bootstrap standard error and percentile limits					2.443 (0.931)	0.707	3.734		
The indirect approach: bootstrap standard error and percentile limits					2.153 (0.910)	0.261	3.074		

Note: Estimates for all fitted models (with one representative for nested models), the model specific BMD estimates, 90% profile likelihood based BMDL and BMDU, corrected AIC and corresponding weight for averaging. Model-averaged BMD estimates based on the direct approach with 90% Wald and percentile bootstrap BMDL and BMDU limits, and based on the indirect approach with 90% percentile bootstrap BMDL and BMDU limits. Estimates for the standard error are shown: using formula (23) and the bootstrap for the direct estimate and only the bootstrap for the indirect estimate. $B = 3,000$ bootstrap samples were generated.

Abbreviations: Inv Exp, inverse exponential; LN, log-normal; LSN, log-skew-normal model.