Published in final edited form as:
J Risk Res. 2010 July ; 13(5): 653–667. doi:10.1080/13669870903551662.

Translational benchmark risk analysis

Walter W. Piegorsch*
Program in Statistics, BIO5 Institute, University of Arizona, Tucson, AZ, USA

Abstract

Translational development – in the sense of translating a mature methodology from one area of application to another, evolving area – is discussed for the use of *benchmark doses* in quantitative risk assessment. Illustrations are presented with traditional applications of the benchmark paradigm in biology and toxicology, and also with risk endpoints that differ from traditional toxicological archetypes. It is seen that the benchmark approach can apply to a diverse spectrum of risk management settings. This suggests a promising future for this important risk-analytic tool. Extensions of the method to a wider variety of applications represent a significant opportunity for enhancing environmental, biomedical, industrial, and socio-economic risk assessments.

Keywords

benchmark concentration; benchmark dose; BMC; BMD; BMDL; environmental biomonitoring; low-dose extrapolation; quantitative risk assessment; risk management; translational research; vulnerability assessment

1. Introduction

Developed in the mid-1980s – but with roots going back much farther and deeper –benchmark analysis for estimating exposure risks to hazardous agents has become a widely accepted technology in quantitative risk assessment. As described by Crump (1984), the method relates an adverse outcome to a quantified dose or exposure, $d \geq 0$, of a hazardous agent via some functional model. It then manipulates components of this model to yield a *benchmark dose* (BMD) of the agent at which a specified *benchmark risk* or *benchmark response* (BMR) is attained. If the exposure is measured as a concentration, one refers to the exposure point as a *benchmark concentration* (BMC). The BMD or BMC is used to arrive at a level of acceptable human or ecological exposure to the agent or to otherwise establish low-exposure guidelines. Risk analysts increasingly use benchmark quantities as the basis for setting occupational exposure limits (OELs) or other 'points of departure' when assessing hazardous stimuli/exposures (Kodell 2005; Kobayashi et al. 2006; Nielsen and Øvrebø 2008). Indeed, both the USA and the Organisation for Economic Cooperation and Development (OECD) provide guidance on BMDs in carcinogen risk assessment (US EPA 2005; OECD 2008), and the use of BMDs or BMCs is growing for quantifying and managing risk with a variety of toxicological endpoints (US General Accounting Office 2001; European Union 2003; OECD 2006). One critical modification is the use of statistical lower confidence limits on the BMD – called benchmark dose (lower) limits or simply BMDLs (Crump 1995) – in order to account for statistical variability of the BMD point estimate.

*piegorsch@math.arizona.edu.

The benchmark paradigm integrates mathematical modeling into the low-dose risk-assessment process, moving away from the so-called no-observed-adverse-effect level (NOAEL) technologies for estimating low exposure doses. (The NOAEL is an older method for estimating virtually safe/low-dose exposure levels, and substantial statistical instabilities have been identified with its use. Most contemporary analysts now recommend against this dated technology [Chapman, Caldwell, and Chapman 1996; Crump 2002a; Kodell 2009].) Critical to the calculation of a BMD is the construction of a mathematical risk function, $R(d)$, to describe how risk changes with increasing levels of dose. In many public health and environmental scenarios, $R(d)$ is further refined into a form of excess risk, such as the *additional risk* $R_A(d) = R(d) - R(0)$ or the *extra risk* $R_E(d) = R_A(d)/\{1 - R(0)\}$. This adjusts for background or spontaneous effects out of the control of the risk regulator (Piegorsch and Bailer 2005, §4.2). Either form may be seen in practice, depending on context.

Under this mathematical modeling approach, the BMD is determined by setting $R(d)$, $R_A(d)$, or $R_E(d)$, as appropriate, equal to the BMR and solving for d. Statistically, this is a form of inverse nonlinear regression, similar to the estimation of an 'effective dose' such as the familiar median effective dose, ED_{50}, in toxicity testing (Piegorsch and Bailer 2005, §4.1). Taken from the unit interval, the BMR is specified in advance of any data acquisition. Calculation of the BMDL can proceed similarly, using, e.g., an upper confidence band on the risk or excess risk from which to base the statistical inversion on to the dose scale (Al-Saidy et al. 2003). Figure 1 from Example 2 illustrates the operations graphically.

The model construction also requires specification of the data's statistical features, which depends on their basic form, e.g., proportions versus continuous measurements. While this is a necessary and critical step in the modeling process, the benchmark paradigm's oft-unrecognized value is its applicability to a variety of statistical models. The method can be applied to any form of outcome and/or dose–response model from which a valid risk function can be defined, estimated from the data, and inverted to calculate BMDs and BMDLs. The goal herein is to highlight the flexibility available under the benchmark approach and illustrate its widening application. It is not intended as a comprehensive review, however, since a number of excellent examples for such already exist (Crump 2002a; Falk Filipsson et al. 2003; Parham and Portier 2005; Sand, Victorin, and Falk Filipsson 2008). Nonetheless, by exploring different forms in which the benchmark paradigm can be applied, readers may discover new and effectual ways to choreograph benchmark analyses for use in quantitative risk assessments.

2. Traditional benchmark analysis with quantal data

The benchmark method is often employed with data in the form of proportions. Therein, the numerators are taken as binomial variates $Y_i \sim \text{Bin}(N_i, R[d_i])$ at each exposure or dose index i, where the proportion denominator N_i is the number of subjects tested, and $R(d_i)$ is used to model the unknown probability that an individual subject will respond at dose $d_i \geq 0$; $i = 1, \ldots, n$. This is called the *quantal response data setting* in dose–response analysis. For such cases, the excess risk function is usually taken to be the extra risk $R_E(d) = \{R(d) - R(0)\}/\{1 - R(0)\}$.

Typically, $R(d)$ is assigned a parametric specification, e.g., the ubiquitous logistic dose–response model $R(d) = 1/(1 + \exp\{-\beta_0 - \beta_1 d\})$, or the similar probit model $R(d) = \Phi(\beta_0 + \beta_1 d)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The unknown β-parameters are estimated from the data; maximum likelihood is a favored estimation approach (Piegorsch and Bailer 2005, §A.4.3). The maximum likelihood estimate (MLE) for the BMD is found by setting the estimated extra risk function, $\hat{R}_E(d)$, equal to the chosen BMR and solving for d. Denote this as $\widehat{\text{BMD}}$. Where needed for clarity, the notation imitates the well-known ED_{50} and subscripts the BMR level at which each quantity is calculated: $\text{BMD}_{100\text{BMR}}$, $\widehat{\text{BMD}}_{100\text{BMR}}$, etc.

Corresponding BMDL(s) are determined using the statistical properties of $\hat{BMD}$; this can involve simple, one-sided Wald confidence limits employing the large-sample characteristics of the MLE, bootstrap confidence bounds, or inversion of simultaneous confidence bands on $R_E(d)$ (cf. Crump and Howe 1985 or Moerbeek, Piersma, and Slob 2004, among many others). In the latter case, the BMDLs inherit the simultaneous confidence features of the bands and can therefore be applied when multiple BMRs are under study (Al-Saidy et al. 2003); see Example 3.

The earliest uses of BMDs and BMDLs for studying low-dose risk with quantal data occurred with laboratory animal toxicity experiments. For example, Crump's original 1984 article described a variety of toxicological applications, involving outcomes such as bacteriological toxicity, developmental anomalies, and *in utero* mortality. Since then, the approach has been applied to many other adverse quantal endpoints, including the important area of mammalian carcinogenesis (Van Landingham et al. 2001; Butterworth, Aylward, and Hays 2007). To illustrate using a contemporary dataset, consider the following example.

2.1. Example 1. Kidney carcinogenesis

The chemical 3-monochloropropane-1,2-diol (3-MCPD) is a contaminant by-product in a variety of foodstuffs, including soy sauces, processed sausages, and breakfast cereals; it can also appear in drinking water after certain types of water treatment. Unfortunately, the chemical is a mammalian nephrotoxin, and ingestion of the agent via such a variety of food and drinking water sources can lead to low-level, chronic population exposures. From a risk-analytic perspective, clearer quantification is required of its potential for producing long-term kidney damage and/or cancer. Towards this end, data were recorded on numbers of female laboratory rats developing renal tumors (adenomas and carcinomas) after exposure to 3-MCPD at a series of increasing doses (reported as mg/kg body weight) in drinking water. Table 1 presents the quantal data, taken from Hwang et al. (2009).

To fit the dose–response data in Table 1, Hwang et al. (2009) chose a probit model, which produced a reasonable fit. The corresponding MLEs for the data in Table 1 yield an estimated risk function of $\hat{R}(d) = \Phi(-2.252 + 0.036d)$. Converting to extra risk and setting BMR to, say, 0.10 – a standard level in toxicology testing – gives $\hat{BMD}_{10} = 28.391$ mg/kg. A 95% Wald lower confidence limit is $BMDL_{10} = 21.305$ mg/kg. (Particulars for the BMDL calculation are detailed in Appendix 1.) Risk regulators can use these values to determine acceptable exposure limits for ingestion of 3-MCPD in any food/drinking water sources contaminated by the chemical.

It is important to note that Hwang et al. (2009) also considered a variety of other dose–response models for their data, many of which operated about as adequately as the probit model featured here. Still, cases can occur where different dose–response models fit a set of quantal data with essentially equivalent quality over the range of observed responses, but may produce substantively different BMDL(s) at very small BMR(s) (Allen et al. 1994; Faustman and Bartell 1997). Contemporary analysts caution against such extreme low-risk/low-dose extrapolations, and often choose BMR(s) in a range that meets or proximates the corresponding lower response rates in the data. In practice, levels of BMR at or between 0.01 and 0.10 are typical (US EPA 2000, §II.B; Kodell 2005). The larger issue of how to adjust or correct for problems of model selection bias/adequacy when the correct form for $R(d)$ is unknown remains an area of active research. Recent work in this direction has led to some intriguing advances, including manipulation of information measures, such as the Akaike Information Criterion (Sand, Falk Filipsson, and Victorin 2002; Moon et al. 2005), and averaging techniques, such as Bayesian model averaging (Bailer, Noble, and Wheeler 2005; Morales et al. 2006), or its frequentist (non-Bayesian) analogs (Faes et al. 2007; Wheeler and Bailer 2007). Much of this research is

still emergent, however, and the statistical science underlying this avenue of BMD estimation remains open for further development.

Increasingly, BMDs are called upon as starting points for toxicological and biomedical risk assessments across a wide variety of endpoints, carcinogenic or otherwise (Gaylor et al. 1999; Kodell 2005). Successful applications with quantal data include adverse developmental endpoints (Krewski, Zhu, and Fung 1999; Zhu, Wang, and Jelsovsky 2007), genetic toxicity (Allen et al. 2005), inhalation toxicity (Fowles, Alexeeff, and Dodge 1999; Gift et al. 2008), and extensions to human epidemiological studies (Yeh, Lee, and Chen 2006). In all the cases, the risk-analytic context may change with the differing endpoint(s), but the fundamentals of the benchmark concept remain essentially unchanged.

3. Benchmark analysis with nonquantal data

In his seminal article, Crump (1984) also described how the BMD could be constructed with nonquantal data, i.e., data not in the form of a proportion *Y*/*N*. In this case, a variety of possible outcomes becomes possible, corresponding to many discrete and continuous statistical distributions (See and Bailer 1998; Banga, Patil, and Taillie 2002; Morales and Ryan 2005). Most common in mammalian toxicological studies is the normal or 'Gaussian' distribution; traditional endpoints, such as weight change and hematological measurements (Slob 2002; Piegorsch et al. 2005b), developmental damage (Razzaghi and Kodell 2004), and neurological outcomes (Kodell, Chen, and Gaylor 1995) often generate normally distributed data, either directly or after logarithmic transformation. Within this context, assume the nonquantal data under study are normal with $Y_i \sim N(\mu[d_i], \sigma^2)$ at each exposure, concentration, or dose index i, where the mean response varies over dose d_i and the variances σ^2 are homogenous but unknown; $i = 1, \ldots, n$. Extensions for heterogeneous variances are also possible (cf. West and Kodell 1993).

Different dose–response models can be chosen to represent $\mu(d)$, depending on the endpoint under study; most common is the linear form $\mu(d) = \beta_0 + \beta_1 d$, or possibly curvilinear extensions such as the parabola $\mu(d) = \beta_0 + \beta_1 d + \beta_2 d^2$ or the general power model $\mu(d) = \beta_0 + \beta_1 d^k$. Modifications that allow for possible threshold effects are easily incorporated into these constructions (Crump 1984).

The use of normal distributions and linear dose–response functions leads to questions on how to define the BMD (Crump 2002b). A formal conceptualization of exposure risk with nonquantal data was given by Gaylor and Slikker (1990) by focusing on the difference between response at exposure dose d and response at the background level $d = 0$. Often called the 'hybrid' benchmark concept, Kodell and West (1993) expanded on this and defined the risk function as:

$$R(d) = P[Y \leq \mu(0) - \kappa\sigma]$$

where $\kappa > 0$ is a constant used to tune $R(d)$ for specific applications. In this form, a response more than κ standard deviations below the control mean is considered adverse. Suggested values include $\kappa = 2$ or $\kappa = 3$. (This formulation assumes that increasing exposure leads to decreases in the response. The risk function is easily modified if the reverse is true.) The fundamental viewpoint embodied in Equation (1) is that BMDs or BMCs for continuous responses should be defined on a probability footing, in similar fashion as that taken with quantal data.

The Kodell–West definition of risk in Equation (1) is often employed in conjunction with an additional risk function, $R_A(d) = R(d) - R(0)$. Under the assumption $Y_i \sim N(\mu[d_i], \sigma^2)$ and

taking, say, a simple linear response function, $\mu(d_i) = \beta_0 + \beta_1 d_i$, the additional risk becomes $R_A(d) = \Phi(-\gamma_1 d - \kappa) - \Phi(-\kappa)$, where $\gamma_1 = \beta_1/\sigma$. If the mean response is quadratic, $\mu(d_i) = \beta_0 + \beta_1 d_i + \beta_2 d_i^2$, the additional risk extends to $R_A(d) = \Phi(-\gamma_1 d - \gamma_2 d^2 - \kappa) - \Phi(-\kappa)$, where $\gamma_j = \beta_j/\sigma$ ($j = 1, 2$).

To estimate the BMD for any given BMR, set $\hat{R}_A(d) = \text{BMR}$ and solve for d. For the linear model with $\mu(d) = \beta_0 + \beta_1 d$, this yields $\widehat{\text{BMD}}_{100\text{BMR}} = -C_{\text{BMR}}/g_1$, where g_1 is an unbiased estimate of γ_1 and $C_{\text{BMR}} = \kappa + \Phi^{-1}\{\text{BMR} + \Phi(-\kappa)\}$ (Piegorsch et al. 2005b). Corresponding BMDLs can be constructed using methods similar to those employed for quantal data; see Appendix 2 for technical details. Extensions to the quadratic case are also possible (Piegorsch et al. 2005a).

The next example illustrates this approach. To emphasize the translational potential of the benchmark paradigm, however, it employs data that differ from the traditional mammalian toxicology setting discussed in Section 2.

3.1. Example 2. Soil ecotoxicity

A growing concern in *ecological risk assessment* is the need to identify damage to ecosystems and to the organisms that populate them. For example, characterization and management of potential ecotoxicological risks to soil (micro)organisms from chemical pollutants is critical for protecting habitat and ecosystem function. Regulators must monitor for contaminants that trigger adverse responses in the soil-inhabiting organisms, and where possible correct for pollutant burdens that drive the risk to unacceptable levels.

For instance, in an environmental biomonitoring study of polycyclic aromatic hydrocarbon (PAH) contamination, Maliszewska-Kordybach et al. (2007) collected samples to study PAH effects on nitrifying bacteria in various soils. Consider their data on light loamy sand contamination, taken in an agricultural region within the south-eastern Polish voivodeship of Lubelskie: measured were reductions in nitrification potential of the soil over varying concentrations of the three-ring PAH phenanthrene (*Phe*), in mg/kg. A decreasing trend in nitrification potential was observed as *Phe* concentrations increased (Maliszewska-Kordybach et al. 2007, Table 3), and the authors found that a simple linear regression model for the mean response provided an adequate fit to their data if they operated with square roots of the *Phe* concentrations, i.e., $d = \sqrt{\textit{Phe}\ \text{conc.}}$ (Final results were then squared for reporting back on the original concentration scale.) The goal was to investigate and summarize the adverse effects of *Phe* soil contamination. To do so Maliszewska-Kordybach et al. employed the well-established median effective concentration, EC_{50}, as their summary toxicity parameter. For a more formal risk assessment of how soil nitrification potential can be affected by *Phe* contamination, however, one can also apply the Kodell–West benchmarking operation described above.

Using the regression estimates reported by Maliszewska-Kordybach et al. (2007), the unbiased point estimate of γ_1 from Appendix 2 calculates as $g_1 = -0.311$. At $\kappa = 3$ the resulting estimator of the additional risk is $\hat{R}_A(d) = \Phi(0.311d - 3.0) - 0.0013$. Setting this equal to any chosen BMR and solving for d, yields the benchmark point estimate. (Since the d-scale here is $\{\textit{Phe}\ \text{conc.}\}^{1/2}$, one should be careful and for purposes of the technical analysis write $\text{BMC}^{1/2}$, although this is admittedly somewhat cumbersome notation.) Following the example of the original study, set BMR = $^1/_2$. This yields $\widehat{\text{BMC}}_{50}^{1/2} = 9.672(\text{mg/kg})^{1/2}$, with 95% lower confidence limit $\text{BMCL}_{50}^{1/2} = 6.196(\text{mg/kg})^{1/2}$; Figure 1 illustrates the operation. Translating back to the original *Phe* scale, one finds $\widehat{\text{BMD}}_{50} = 93.543$ mg/kg and $\text{BMCL}_{50} = 38.396$ mg/kg.

These benchmark values allow for ecotoxicological monitoring of *Phe* soil contamination and for evaluation of the ecological status of the affected soil. Measured concentrations found at or above these limits indicate substantive adverse effects of the chemical on nitrification potential. This can motivate soil quality managers or environmental regulators to institute quarantines, avoid larger environmental contamination, and/or encourage remediation efforts in support of the affected ecosystem.

As Example 2 highlights, while the biological and toxicological context for construction of BMDs, BMCs, etc. may vary, the fundamental benchmarking concept remains unchanged. Indeed, ecotoxicological benchmark analyses with nonquantal data are growing (Wu et al. 2006; Suski et al. 2008), and barely keep pace with applications in more traditional, nonquantal, toxicity studies (Slob 2002) or with epidemiological investigations on nonquantal human endpoints (Bailer et al. 1997; Budtz-Jørgensen, Keiding, and Grandjean 2001; Morales and Ryan 2005). Expansion has even progressed into rapidly developing areas of biological research such as genomics/gene-expression analysis (Yang, Allen, and Thomas 2007). Opportunities therefore exist for much greater use of the benchmark paradigm, limited only by the creativity of the risk analyst. The next section describes a more divergent possibility, with an eye towards giving benchmark analysis greater translational standing in risk management and practice.

4. Other forms of translational benchmark analysis

As suggested above, an under-recognized feature of the benchmark approach is its applicability to a variety of risk-analytic scenarios. The previous examples illustrate how broadly the method can be applied within biology and toxicology, but it takes surprisingly little effort to extend it into distinctly separate areas of risk assessment.

At its core, the mathematical models that support the method can be manipulated across an assortment of data scenarios. Certainly, when the context of a particular risk analysis changes – say, from crude lethality to mammalian carcinogenesis, to ecotoxicity – the observed data can move from proportions to continuous measurements, etc. The conceptual features are essentially unaffected, however, since the basic statistical methodology remains fungible across different definitions of the adverse event. This facilitates transfer of the benchmark technology to any applicable form of adverse outcome and/or dose–response model, translating the BMDs and BMDLs to many wide-ranging problems in quantitative risk assessment.

To illustrate, the following example highlights the method in a setting far afield from the traditional toxicological arena: socio-geographic hazard analysis.

4.1 Example 3. Place-based terrorism vulnerability

Hazard/vulnerability indexing for populations, localities, ecosystems, and other entities at risk to adverse events is an emergent endeavor in twenty-first-century risk assessment, which has only recently seen substantial development (Peng, Schoenberg, and Woods 2005; Tsuzuki 2006; Mazzorana et al. 2009). For instance, a recent study on urban risk/vulnerability to terrorist attacks (Piegorsch, Cutter, and Hardisty 2007) employed a quantitative index to characterize how terrorism vulnerability changes across different urban locations. The vulnerability data indicated whether any of the 132 largest cities in the USA experienced a terrorist attack over the 35-year period 1970–2004 in which casualties (human deaths or injuries) occurred. Thus, the observed response was a simple binary indicator: $Y_i = 1$ if casualties had been recorded in that locale during this time period, or 0 if not. This was compared to an index – called the place vulnerability index (PVI) – that combined the locality's spectrum of socio-economic, geo-physical, and built-environment vulnerability to hazardous events. (The study also considered other aspects of terrorism vulnerability; see the 2007 article for details and also for the complete

132 location dataset.) The PVI had no units, and was standardized to range between roughly −4 and 4. Its scale was uniquely defined, however, and as such the index could be viewed as a ‘dose’ metric for urban vulnerability: as the PVI increased, urban vulnerability to terrorism casualties also appeared to increase.

In fact, the association between urban terrorism causalities and the PVI can be modeled effectively using common dose–response functions. It was seen that a form of quantal response relationship known as the complementary log–log (CLL) model could describe the terrorism ‘dose response’: if d_i = PVI for the ith locality, then the probability of observing $Y_i = 1$ (i.e., terrorist casualties) at that locality was taken as $R(d_i) = 1 - \exp\{-\exp[\beta_0 + \beta_1 d_i]\}$.

Using the urban casualty data, the goal was to identify values of PVI past which a locality’s vulnerability to terrorism casualties exceeded reasonable levels. This could be addressed via application of the benchmark approach. The risk assessor defines one or more BMR(s) as the distinguishing points for an unacceptable level of risk, sets the risk function equal to the BMR, and solves for the vulnerability index d = PVI. This produces a *benchmark index*, BMI, at which the risk exceeds the chosen level of BMR. Note that, because the PVI is a unitless measure, no adjustment is needed for background or spontaneous effects. Thus, one operates directly on the risk function: set $R(d) = \text{BMR}$ and solve for d. Under the CLL model this produces $\text{BMI}_{100\text{BMR}} = (\log\{-\log[1 - \text{BMR}]\} - \beta_0)/\beta_1$.

Since these terrorism casualty data are binary, a simplified binomial model, $Y_i \sim \text{Bin}(1, R[d_i])$, is valid, and MLEs for the unknown β-parameters can be determined using similar techniques as those in Appendix 1; see Piegorsch, Cutter, and Hardisty (2007) for details. Doing so for the casualty vulnerability data yields an estimated risk function of $\hat{R}(d) = 1 - \exp\{-\exp[-1.255 + 0.474d]\}$. Inverting the risk produces $\widehat{\text{BMI}}_{100\text{BMR}} = (\log\{-\log[1 - \text{BMR}]\} + 1.255)/0.474$. For example, the standard quartiles $^1/_4$, $^1/_2$, $^3/_4$ represent symmetric BMR cut-points that characterize levels of increasing vulnerability. For the 132 cities data, this gives $\widehat{\text{BMI}}_{25} = 0.018$, $\widehat{\text{BMI}}_{50} = 1.873$, and $\widehat{\text{BMI}}_{75} = 3.335$.

For the corresponding benchmark index (lower) limits (BMILs), adjustment is necessary here for the use of multiple BMRs, which can be achieved by inverting a 95% simultaneous upper confidence band on the extra risk (Al-Saidy et al. 2003). This yields $\text{BMIL}_{25} = -1.820$, $\text{BMIL}_{50} = 1.001$, and $\text{BMIL}_{75} = 1.986$. Figure 2 illustrates the operations, where the similarities to Figure 1 are evident.

Further mimicking the analysis in the original article, one could return to these data and ask which of the 132 cities lie above, say, the upper risk quartile’s BMIL_{75} of 1.986? Doing so identifies 5 of the 132 urban centers with extreme-level PVIs: New York, NY–Newark, NJ (PVI = 2.154), Norfolk–Chesapeake–Newport News–Virginia Beach, VA (PVI = 2.326), Charleston, SC (PVI = 2.543), Baton Rouge, LA (PVI = 3.016), and New Orleans, LA (PVI = 3.119). Intriguingly, all are port cities, hinting at an important marker for increased vulnerability to terrorist events. (Admittedly, a number of other port cities in the list, e.g., San Diego, CA, scored much lower PVIs. Nonetheless, these results give suggestions into the roots of urban vulnerability to terrorism.)

This untraditional application of benchmark analysis helps focus attention on those urban centers whose PVIs exceed defined benchmarks such as BMIL_{75}. From it, city officials in, e.g., Charleston, SC or Norfolk, VA may wish to consider new or updated forms of coastal anti-terrorist protection. The analysis may also prompt increases in funding allocation(s) or other heightened efforts to connect urban risk management programs with terrorism vulnerability assessments (Kunreuther 2002).

5. Discussion

As these various examples demonstrate, the benchmark paradigm's ability to extend past its roots in toxicity testing and its potential for further implementation are quite rich. One message has pervaded throughout: even when the particular risk-analytic context changes – in the nature of the subject matter problem and/or the specifics of the statistical model – the ability of the benchmark approach to accommodate these differences is unaffected. While the context changes, the concept retains its flexibility. This was recognized by many of the authors cited herein as they employed the method to find BMDs, BMCs, BMIs, etc. More can be done, however, and as Barnes et al. (1995) concluded in a report from a 1994 Benchmark Dose Workshop, 'to achieve general acceptance of the BMD approach, it will have to be applied to a variety of endpoints'. A marked impact on risk assessment practice will then ensue. As is true in many areas of scientific endeavor, by studying familiar processes in unfamiliar settings we gain a better grasp of both the familiar processes and the new settings.

Acknowledgments

Thanks are due to an anonymous referee for her/his helpful comments, and Dr Ralph L. Kodell for his expert input on benchmark analysis and for his supportive suggestions on various components of this manuscript. This research was supported by grant #RD-83241902 from the US Environmental Protection Agency and by grant #R21-ES016791 from the US National Institute of Environmental Health Sciences. Its contents are solely the responsibility of the author and do not necessarily reflect the official views of these agencies.

References

Akahira M. A higher-order approximation to a percentage point of the noncentral t-distribution. Communications in Statistics: Simulation and Computation 1995;24:595–605.

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM. Dose–response assessment for developmental toxicity. II: Comparison of generic benchmark dose estimates with no observed adverse effect levels. Fundamental and Applied Toxicology 1994;23:487–95. [PubMed: 7867900]

Allen B, Zeiger E, Lawrence G, Friedman M, Shipp A. Dose–response modeling of in vivo genotoxicity data for use in risk assessment: Some approaches illustrated by an analysis of acrylamide. Regulatory Toxicology and Pharmacology 2005;41:6–27. [PubMed: 15649824]

Al-Saidy OM, Piegorsch WW, West RW, Nitcheva DK. Confidence bands for low-dose risk estimation with quantal response data. Biometrics 2003;59:1056–62. [PubMed: 14969485]

Bailer AJ, Noble RB, Wheeler MW. Model uncertainty and risk estimation for experimental studies of quantal responses. Risk Analysis 2005;25:291–9. [PubMed: 15876205]

Bailer AJ, Stayner LT, Smith RJ, Kuempel ED, Prince MM. Estimating benchmark concentrations and other noncancer endpoints in epidemiology studies. Risk Analysis 1997;17:771–80. [PubMed: 9463931]

Banga SJ, Patil GP, Taillie C. Continuous dose–response modeling and risk analysis with the gamma and reciprocal gamma distributions. Environmental and Ecological Statistics 2002;9:273–93.

Barnes DG, Daston GP, Evans JS, Jarabek AM, Kavlock RJ, Kimmel CA, Park C, Spitzer HL. Benchmark dose workshop: Criteria for use of a benchmark dose to estimate a reference dose. Regulatory Toxicology and Pharmacology 1995;21:296–306. [PubMed: 7644719]

Budtz-Jørgensen E, Keiding N, Grandjean P. Benchmark dose calculation from epidemiological data. Biometrics 2001;57:698–706. [PubMed: 11550917]

Butterworth BE, Aylward LL, Hays SM. A mechanism-based cancer risk assessment for 1,4-dichlorobenzene. Regulatory Toxicology and Pharmacology 2007;49:138–48. [PubMed: 17688981]

Chapman PM, Caldwell RS, Chapman PF. A warning: NOECs are inappropriate for regulatory use. Environmental Toxicology and Chemistry 1996;15:77–9.

Crump KS. A new method for determining allowable daily intake. Fundamental and Applied Toxicology 1984;4:854–71. [PubMed: 6510615]

Crump KS. Calculation of benchmark doses from continuous data. Risk Analysis 1995;15:79–89.

Crump, KS. Benchmark analysis. In: El-Shaarawi, AH.; Piegorsch, WW., editors. Encyclopedia of environmetrics. Vol. 1. Chichester: John Wiley; 2002a. p. 163-70.

Crump KS. Critical issues in benchmark calculations from continuous data. Critical Reviews in Toxicology 2002b;32:133–53. [PubMed: 12071571]

Crump, KS.; Howe, R. A review of methods for calculating confidence limits in low dose extrapolation. In: Clayson, DB.; Krewski, D.; Munro, I., editors. Toxicological risk assessment. Vol. 1, Biological and statistical criteria. Boca Raton, FL: CRC Press; 1985. p. 187-203.

European Union. Technical Report No. EUR 20418 EN/1-4. European Chemicals Bureau (ECB); Ispra, Italy: 2003. Technical Guidance Document (TGD) on risk assessment of chemical substances following European Regulations and Directives, Parts I–IV.

Faes C, Aerts M, Geys H, Molenberghs G. Model averaging using fractional polynomials to estimate a safe level of exposure. Risk Analysis 2007;27:111–23. [PubMed: 17362404]

Falk Filipsson A, Sand S, Nilsson J, Victorin K. The benchmark dose method: Review of available models, and recommendations for application in health risk assessment. Critical Reviews in Toxicology 2003;33:505–42. [PubMed: 14594105]

Faustman EM, Bartell SM. Review of noncancer risk assessment: Applications of benchmark dose methods. Human and Ecological Risk Assessment 1997;3:893–920.

Fowles JR, Alexeeff GV, Dodge D. The use of benchmark dose methodology with acute inhalation lethality data. Regulatory Toxicology and Pharmacology 1999;29:262–78. [PubMed: 10388612]

Gaylor DW, Kodell RL, Chen JJ, Krewski D. A unified approach to risk assessment for cancer and noncancer endpoints based on benchmark doses and uncertainty/safety factors. Regulatory Toxicology and Pharmacology 1999;29:151–7. [PubMed: 10341145]

Gaylor DW, Slikker WL. Risk assessment for neurotoxic effects. NeuroToxicology 1990;11:211–8. [PubMed: 1700346]

Gift JS, McGaughy R, Singh DV, Sonawane B. Health assessment of phosgene: Approaches for derivation of reference concentration. Regulatory Toxicology and Pharmacology 2008;51:98–107. [PubMed: 18440110]

Hwang M, Yoon E, Kim J, Jang DD, Yoo TM. Toxicity value for 3-monochlo-ropropane-1,2-diol using a benchmark dose methodology. Regulatory Toxicology and Pharmacology 2009;53:102–6. [PubMed: 19133308]

Kobayashi E, Suwazono Y, Uetani M, Inaba T, Oishi M, Kido T, Nishijo M, Nakagawa H, Nogawa K. Estimation of benchmark dose for renal dysfunction in a cadmium non-polluted area in Japan. Journal of Applied Toxicology 2006;26:351–5. [PubMed: 16791912]

Kodell RL. Managing uncertainty in health risk assessment. International Journal of Risk Assessment and Management 2005;5:193–205.

Kodell RL. Replace the NOAEL and LOAEL with the $BMDL_{01}$ and $BMDL_{10}$. Environmental and Ecological Statistics 2009;16:3–12.

Kodell RL, Chen JJ, Gaylor DW. Neurotoxicity modeling for risk assessment. Regulatory Toxicology and Pharmacology 1995;22:24–9. [PubMed: 7494899]

Kodell RL, West RW. Upper confidence intervals on excess risk for quantitative responses. Risk Analysis 1993;13:177–82. [PubMed: 8502790]

Krewski D, Zhu Y, Fung KY. Benchmark doses for developmental toxicants. Inhalation Toxicology 1999;11:579–92. [PubMed: 11203000]

Kunreuther H. Risk analysis and risk management in an uncertain world. Risk Analysis 2002;22:655–64. [PubMed: 12224739]

Kutner, MH.; Nachtsheim, CJ.; Neter, J. Applied linear regression models. 4. Boston, MA: McGraw-Hill Irwin; 2004.

Maliszewska-Kordybach B, Klirnkowicz-Pawlas A, Smreczak B, Janusauskaite D. Ecotoxic effect of phenanthrene on nitrifying bacteria in soils of different properties. Journal of Environmental Quality 2007;36:1635–45. [PubMed: 17940263]

Mazzorana B, Zischg A, Largiader A, Hübl J. Hazard index maps for woody material recruitment and transport in alpine catchments. Natural Hazards and Earth System Sciences 2009;9:197–209.

Moerbeek M, Piersma AH, Slob W. A comparison of three methods for calculating confidence intervals for the benchmark dose. Risk Analysis 2004;24:31–40. [PubMed: 15027998]

Moon H, Kim HJ, Chen JJ, Kodell RL. Model averaging using the Kullback information criterion in estimating effective doses for microbial infection and illness. Risk Analysis 2005;25:1147–59. [PubMed: 16297221]

Morales KH, Ibrahim JG, Chen CJ, Ryan LM. Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. Journal of the American Statistical Association 2006;101:9–17.

Morales KH, Ryan LM. Benchmark dose estimation based on epidemiologic cohort data. Environmetrics 2005;16:435–47.

Nielsen GD, Øvrebø S. Background, approaches and recent trends for setting health-based occupational exposure limits: A minireview. Regulatory Toxicology and Pharmacology 2008;51:253–69. [PubMed: 18502550]

OECD. Document No. 54. Paris: Environment Directorate, Organisation for Economic Co-Operation and Development; 2006. Current approaches in the statistical analysis of ecotoxicity data: A guidance to application, Series on testing and assessment.

OECD. Draft guidance document on the performance of chronic toxicity and carcinogenicity studies, supporting TG 451, 452 and 453. Paris: Organisation for Economic Co-Operation and Development; 2008.

Parham, F.; Portier, CJ. Benchmark dose approach. In: Edler, L.; Kitsos, C., editors. Recent advances in quantitative methods in cancer and human health risk assessment. Chichester: John Wiley; 2005. p. 239-54.

Peng RD, Schoenberg FP, Woods JA. A space-time conditional intensity model for evaluating a wildfire hazard index. Journal of the American Statistical Association 2005;100:26–35.

Piegorsch, WW.; Bailer, AJ. Analyzing environmental data. Chichester: John Wiley; 2005.

Piegorsch WW, Cutter SL, Hardisty F. Benchmark analysis for quantifying urban vulnerability to terrorist incidents. Risk Analysis 2007;27:1411–25. [PubMed: 18093043]

Piegorsch WW, West RW, Pan W, Kodell RL. Low-dose risk estimation via simultaneous statistical inferences. Journal of the Royal Statistical Society, series C (Applied Statistics) 2005a;54:245–58.

Piegorsch WW, West RW, Pan W, Kodell RL. Simultaneous confidence bounds for low-dose risk assessment with non-quantal data. Journal of Biopharmaceutical Statistics 2005b;15:17–31. [PubMed: 15702602]

R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2009.

Razzaghi M, Kodell RL. Quantitative risk assessment for developmental neurotoxic effects. Risk Analysis 2004;24:1673–81. [PubMed: 15660620]

Sand S, Falk Filipsson A, Victorin K. Evaluation of the benchmark dose method for dichotomous data: Model dependence and model selection. Regulatory Toxicology and Pharmacology 2002;36:184–97. [PubMed: 12460753]

Sand S, Victorin K, Falk Filipsson A. The current state of knowledge on the use of the benchmark dose concept in risk assessment. Journal of Applied Toxicology 2008;28:405–21. [PubMed: 17879232]

See K, Bailer AJ. Added risk and inverse estimation for count responses in reproductive aquatic toxicology studies. Biometrics 1998;54:67–73. [PubMed: 9544508]

Slob W. Dose–response modeling of continuous endpoints. Toxicological Sciences 2002;66:298–312. [PubMed: 11896297]

Suski JG, Salice C, Houpt JT, Bazar MA, Talent LG. Dose-related effects following oral exposure of 2,4-dinitrotoluene on the western fence lizard, *Sceloporus occidentalis*. Environmental Toxicology and Chemistry 2008;27:352–9. [PubMed: 18348623]

Tsuzuki Y. An index directly indicates land-based pollutant load contributions of domestic wastewater to the water pollution and its application. Science of the Total Environment 2006;370:425–40. [PubMed: 16916535]

US EPA. External Review Draft No. EPA/630/R-00/001. Washington, DC: US Environmental Protection Agency; 2000. Benchmark dose technical guidance document.

US EPA. Technical Report No. EPA/630/P-03/001F. Washington, DC: US Environmental Protection Agency; 2005. Guidelines for carcinogen risk assessment.

US General Accounting Office. Report to Congressional Requesters No. GAO-01-810. Washington, DC: US General Accounting Office; 2001. Chemical risk assessment: Selected federal agencies' procedures, assumptions, and policies.

Van Landingham CB, Allen BC, Shipp AM, Crump KS. Comparison of the EU T25 single point estimate method with benchmark dose response modeling for estimating potency of carcinogens. Risk Analysis 2001;21:641–56. [PubMed: 11726018]

West RW, Kodell RL. Statistical methods of risk assessment for continuous variables. Communications in Statistics: Theory and Methods 1993;22:3363–76.

Wheeler MW, Bailer AJ. Properties of model-averaged BMDLs: A study of model averaging in dichotomous response risk estimation. Risk Analysis 2007;27:659–70. [PubMed: 17640214]

Wu Y, Piegorsch WW, West RW, Tang D, Petkewich MO, Pan W. Multiplicity-adjusted inferences in risk assessment: Benchmark analysis with continuous response data. Environmental and Ecological Statistics 2006;13:125–41.

Yang L, Allen BC, Thomas RS. BMDExpress: A software tool for the benchmark dose analyses of genomic data. BMC Genomics 2007;8 Article 387.

Yeh YY, Lee HK, Chen WM. Cancer risk analysis and assessment of trihalomethanes in drinking water. Stochastic Environmental Research and Risk Assessment 2006;21:1–13.

Zhu Y, Wang T, Jelsovsky JZH. Bootstrap estimation of benchmark doses and confidence limits with clustered quantal data. Risk Analysis 2007;27:447–65. [PubMed: 17511711]

Appendix 1. Calculating BMDs and BMDLs under a probit dose–response assumption

Suppose data are taken as binomial variates, $Y_i \sim \text{Bin}(N_i, \text{R}[d_i])$, at each exposure or dose index i, where the proportion denominator N_i is the number of subjects tested and $\text{R}(d_i)$ is the unknown risk function; $i = 1, \ldots, n$. Set fixed dose levels $d_i \geq 0$, and assume the risk satisfies a simple probit dose response: $\text{R}(d) = \Phi(\beta_0 + \beta_1 d)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Estimation of the parameter vector $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ proceeds via maximum likelihood: maximize the log-likelihood function $\ell(b) = \Omega_i + \sum_{i=1}^{n} Y_i \log\{\Phi(\beta_0 + \beta_1 d_i)/[1 - \Phi(\beta_0 + \beta_1 d_i)]\} + \sum_{i=1}^{n} N_i \log\{1 - \Phi(\beta_0 + \beta_1 d_i)\}$, where Ω_i is a constant not dependent upon β. Unfortunately, a closed-form solution is not possible in this case and hence calculation proceeds via computer iteration. A number of computer packages can perform the operations, producing MLEs b_0 for β_0 and b_1 for β_1; herein, the R statistical language was used (R Development Core Team 2009). The resulting MLE for the risk function is $\hat{\text{R}}(d) = \Phi(b_0 + b_1 d)$.

To estimate the BMD, find the extra risk as $\text{R}_E(d) = \{\text{R}(d) - \text{R}(0)\}/\{1 - \text{R}(0)\} = \{\Phi(\beta_0 + \beta_1 d) - \Phi(\beta_0)\}/\{1 - \Phi(\beta_0)\}$. The corresponding MLE is $\hat{\text{R}}_E(d) = \{\Phi(b_0 + b_1 d) - \Phi(b_0)\}/\{1 - \Phi(b_0)\}$. Set this equal to the pre-specified value(s) of BMR and solve for d:

$$\frac{\Phi(b_0 + b_1 d) - \Phi(b_0)}{1 - \Phi(b_0)} = \text{BMR} \Rightarrow \widehat{\text{BMD}}_{100\text{BMR}} = \frac{Q_{\text{BMR}} - b_0}{b_1}$$

where $Q_{\text{BMR}} = \Phi^{-1}\{\text{BMR}[1 - \Phi(b_0)] + \Phi(b_0)\}$.

To compute a BMDL, a number of possible formulations exist. The simplest, and the one employed above in Example 1, is the so-called Wald $100(1 - \alpha)\%$ lower limit

$$\mathrm{BMDL}_{100\mathrm{BMR}} = \widehat{\mathrm{BMD}}_{100\mathrm{BMR}} - z_{\alpha}\,\mathrm{se}[\widehat{\mathrm{BMD}}_{100\mathrm{BMR}}]$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$, and $se[\widehat{\mathrm{BMD}}_{100\mathrm{BMR}}]$ is the standard error of $\widehat{\mathrm{BMD}}_{100\mathrm{BMR}}$. The standard error can be approximated via the large-sample features of the ML approach (Piegorsch and Bailer 2005, §A.5). For the probit model assumed here, the squared standard error is

$$se^2\left[\widehat{\mathrm{BMD}}_{100\mathrm{BMR}}\right] \approx \frac{1}{\mathrm{b}_1^2}\left[(1-\mathrm{BMR})\exp\left\{\left(\mathrm{Q}_{\mathrm{BMR}}^2 - \mathrm{b}_0^2\right)/2\right\} - 1\right]^2 se^2[\mathrm{b}_0]$$
$$-\frac{2\widehat{\mathrm{BMD}}_{100\mathrm{BMR}}}{\mathrm{b}_1^2}\left[(1-\mathrm{BMR})\exp\left\{\left(\mathrm{Q}_{\mathrm{BMR}}^2 - \mathrm{b}_0^2\right)/2\right\} - 1\right] cov[\mathrm{b}_0, \mathrm{b}_1] + \frac{1}{\mathrm{b}_1^2}\left[\widehat{\mathrm{BMD}}_{100\mathrm{BMR}}\right]^2 se^2[\mathrm{b}_1]$$

where $se[\mathrm{b}_j]$ are the standard errors of b_j ($j = 0, 1$), and $cov[\mathrm{b}_0,\mathrm{b}_1]$ is their estimated covariance. These latter quantities are readily available from most software routines for performing a probit regression fit.

Appendix 2. Calculating BMDs and BMDLs under a simple linear model assumption

Suppose data are taken as normal or 'Gaussian' variates, $Y_i \sim \mathrm{N}(\mu[d_i], \sigma^2)$, at each exposure, concentration, or dose $d_i \geq 0$, where the mean function is the simple linear form $\mu(d_i) = \beta_0 + \beta_1 d_i$ and σ^2 is a constant, unknown variance term; $i = 1, \ldots, n$. Estimation of the β-parameters and σ^2 proceeds via standard least-squares (LS) techniques (Kutner, Nachtsheim, and Neter 2004). Also, let $S_{dd} = \sum_{i=1}^{n}(d_i - \bar{d})$.

Under the Kodell–West construction in Equation (1), the estimated additional risk is $\hat{\mathrm{R}}_{\mathrm{A}}(d) = \Phi(-g_1 d - \kappa) - \Phi(-\kappa)$, where $\kappa > 0$ is a pre-specified tuning parameter and g_1 is an unbiased estimator of the ratio $\gamma_1 = \beta_1/\sigma$. The latter term is shown to be $g_1 = \xi_{\mathrm{n}}\mathrm{b}_1/\sqrt{\mathrm{MSE}}$, with

$$\xi_{\mathrm{n}} = \frac{\Gamma\left(\frac{\mathrm{n}-2}{2}\right)}{\Gamma\left(\frac{\mathrm{n}-3}{2}\right)}\sqrt{\frac{2}{\mathrm{n}-2}}$$

and where b_1 is the LS estimator of β_1, $\sqrt{\mathrm{MSE}}$ is the root mean squared error, and $\Gamma(\cdot)$ is the gamma function. Notice that this requires $n \geq 4$.

To estimate the BMD for any given BMR, set $\Phi(-g_1 d - \kappa) - \Phi(-\kappa) = \mathrm{BMR}$ and solve for d. This yields $\widehat{\mathrm{BMD}}_{100\mathrm{BMR}} = -\mathrm{C}_{\mathrm{BMR}}/g_1$, where $\mathrm{C}_{\mathrm{BMR}} = \kappa + \Phi^{-1}\{\mathrm{BMR} + \Phi(-\kappa)\}$. To compute a BMDL, find any $100(1-\alpha)\%$ lower bound G_α that satisfies $\mathrm{P}[\mathrm{G}_\alpha < \gamma_1] = 1 - \alpha$. Then, a $100(1-\alpha)\%$ lower bound on the $\mathrm{BMD}_{100\mathrm{BMR}}$ is $\mathrm{BMDL}_{100\mathrm{BMR}} = -\mathrm{C}_{\mathrm{BMR}}/\mathrm{G}_\alpha$.

Piegorsch et al. (2005b) studied different lower bounds for G_α, and favored a choice based on a Cornish–Fisher expansion for the random variable $T_1 = g_1 S_{dd}^{1/2}/\xi_n$ given by Akahira (1995). This led to:

$$G_\alpha = \frac{\lambda_n T_1 - z_\alpha \sqrt{1 + T_1^2\left(1 - \lambda_n^2\right)} + \frac{T_1^3\left(z_\alpha^2 - 1\right)}{24(n-2)^2\left\{1 + T_1^2\left(1 - \lambda_n^2\right)\right\}}\left\{1 + \frac{1}{4(n-2)}\right\}}{\sqrt{S_{dd}}}$$

where

$$\lambda_n = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\sqrt{\frac{2}{n-2}}$$

Piegorsch et al. (2005b) also showed that for positive doses $d > 0$, the BMDL can be extended into *simultaneous* $100(1 - \alpha)\%$ lower bounds on BMD for any BMR: one first uses G_α to construct a simultaneous upper confidence band on $R_A(d)$ satisfying $P[R_A(d) < \Phi(-G_\alpha d - \kappa) - \Phi(-\kappa), \forall\, d > 0] = 1 - \alpha$, and then inverts the band to find the BMDL at the chosen BMR. Figure 1 in Example 2, above, illustrates the operation.

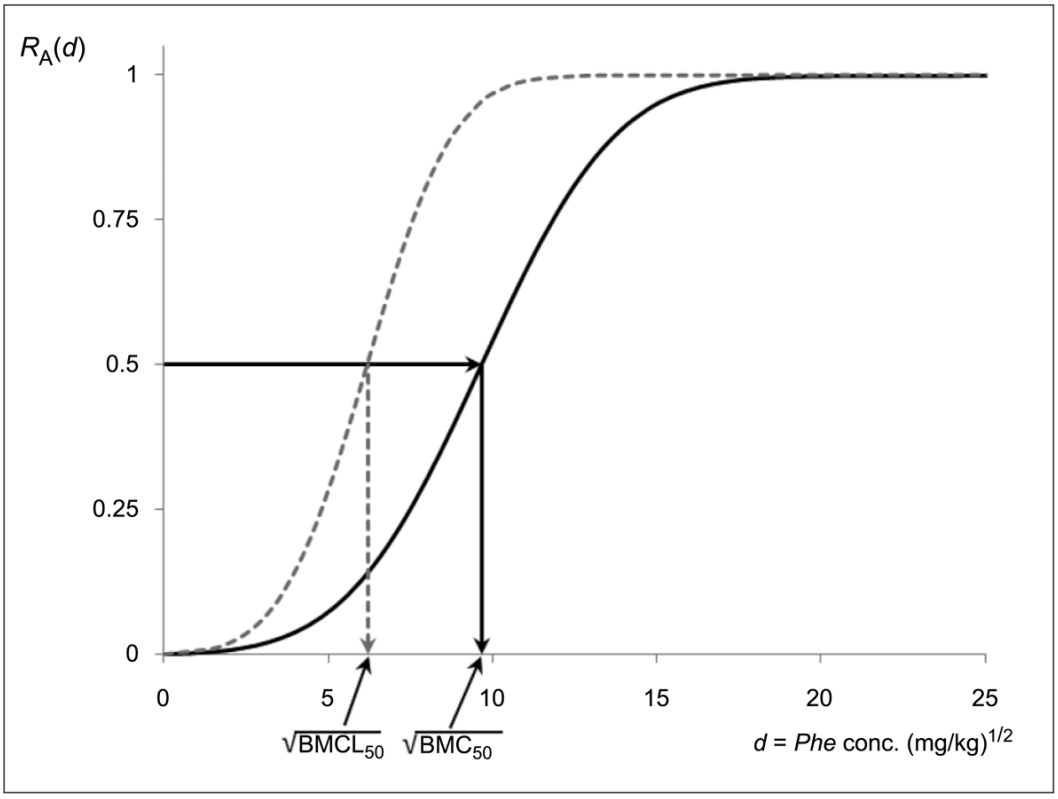

Figure 1.

$BMC_{50}^{1/2}$ and 95% $BMCL_{50}^{1/2}$ (at BMR = 0.50) for the soil ecotoxicity data.
Note: Dark curve is estimated additional risk function $\hat{R}_A(d)$; dashed curve is 95% upper confidence band on the extra risk; cf. Appendix 2.

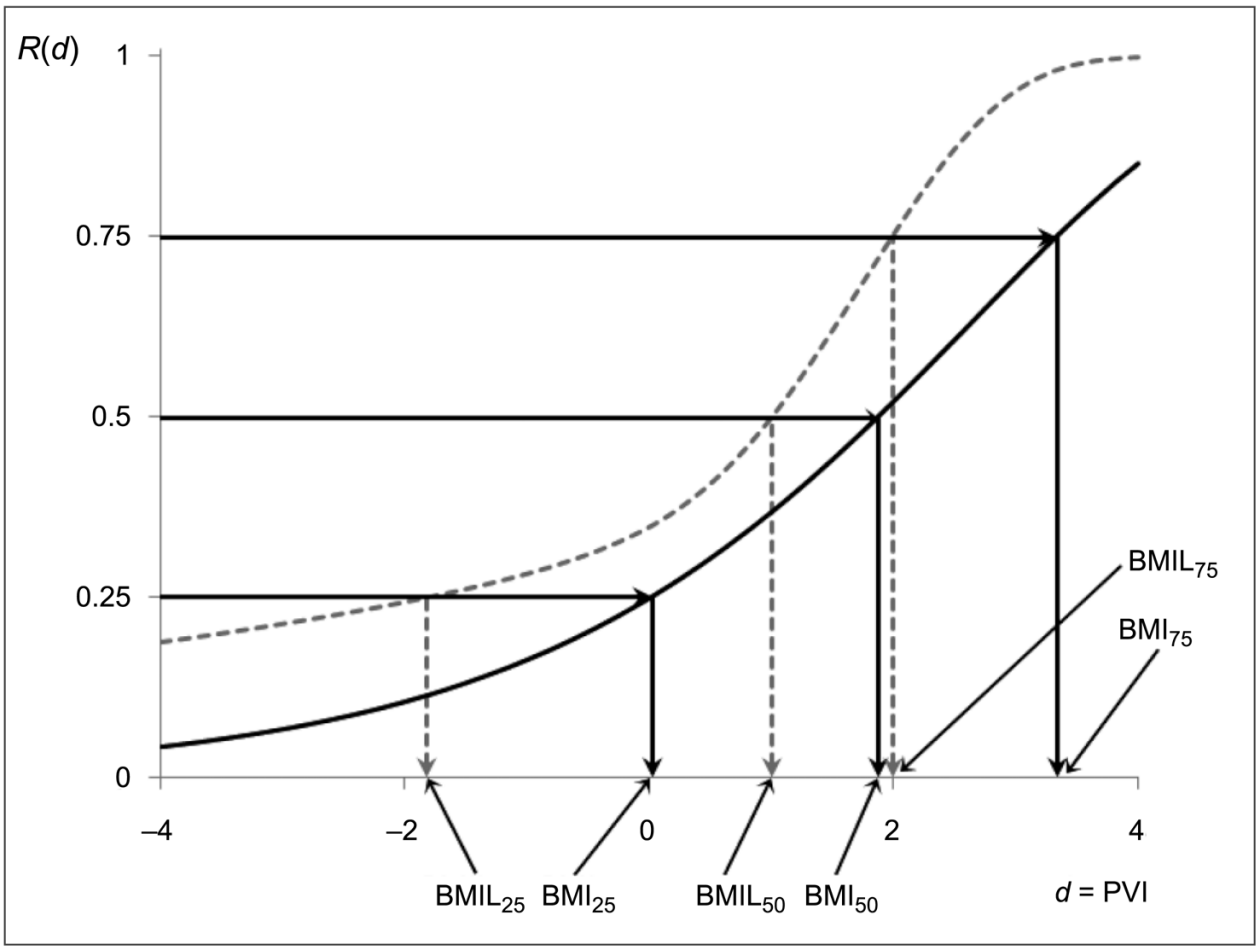

Figure 2.
BMIs and simultaneous 95% BMILs (at BMR = 0.25, 0.50, 0.75) for the terrorism vulnerability data. Note: Dark curve is estimated risk function $\hat{R}(d)$; dashed curve is 95% simultaneous upper band on the extra risk.

Table 1

Kidney adenomas and carcinomas in female laboratory rats after exposure to the food/water contaminant 3-MCPD.

Dose, *d* (mg/kg)	Number with cancer, *Y*	Number exposed, *N*
0	1	50
2.68	0	50
10.34	2	50
37.03	9	50

Note: Kidney carcinogenesis data from Hwang et al. (2009).