

6-2011

GENERIC HOCKEY-STICK MODEL FOR ESTIMATING BENCHMARK DOSE AND POTENCY: PERFORMANCE RELATIVE TO BMDS AND APPLICATION TO ANTHRAQUINONE

Kenneth T Bogen

Exponent Health Sciences, Oakland, CA

Follow this and additional works at: https://scholarworks.umass.edu/dose_response

Recommended Citation

Bogen, Kenneth T (2011) "GENERIC HOCKEY-STICK MODEL FOR ESTIMATING BENCHMARK DOSE AND POTENCY: PERFORMANCE RELATIVE TO BMDS AND APPLICATION TO ANTHRAQUINONE," *Dose-Response: An International Journal*: Vol. 9 : Iss. 2 , Article 4.

Available at: https://scholarworks.umass.edu/dose_response/vol9/iss2/4

This Article is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Dose-Response: An International Journal by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

GENERIC HOCKEY-STICK MODEL FOR ESTIMATING BENCHMARK DOSE AND POTENCY: PERFORMANCE RELATIVE TO BMDS AND APPLICATION TO ANTHRAQUINONE

Kenneth T. Bogen □ Exponent Health Sciences, Oakland, California

□ Benchmark Dose Model software (BMDS), developed by the U.S. Environmental Protection Agency, involves a growing suite of models and decision rules now widely applied to assess noncancer and cancer risk, yet its statistical performance has never been examined systematically. As typically applied, BMDS also ignores the possibility of reduced risk at low doses (“hormesis”). A simpler, proposed Generic Hockey-Stick (GHS) model also estimates benchmark dose and potency, and additionally characterizes and tests objectively for hormetic trend. Using 100 simulated dichotomous-data sets (5 dose groups, 50 animals/group), sampled from each of seven risk functions, GHS estimators performed about as well or better than BMDS estimators, and a surprising observation was that BMDS mis-specified all of six non-hormetic sampled risk functions most or all of the time. When applied to data on rodent tumors induced by the genotoxic chemical carcinogen anthraquinone (AQ), the GHS model yielded significantly negative estimates of net potency exhibited by the combined rodent data, suggesting that—consistent with the anti-leukemogenic properties of AQ and structurally similar quinones—environmental AQ exposures do not likely increase net cancer risk. In addition to its simplicity and flexibility, the GHS approach offers a unified, consistent approach to quantifying environmental chemical risk.

Key Words: Bootstrap, modeling, Monte Carlo, toxicity risk characterization

1. INTRODUCTION

Dose-response assessment is an essential step in characterizing the extent to which environmental contaminants pose human health risks (NRC, 1983, 1994). Benchmark Dose Model software (BMDS) and related procedures (EPA 2000a, 2010a b) comprise a suite of dose-response models and a related set of quasi-statistical and decision rules developed by the U.S. Environmental Protection Agency (EPA) that are increasingly broadly applied as an integral component of regulatory noncancer and cancer risk assessment for environmental chemicals. In particular, this approach is used to estimate “benchmark dose” (BMD) and corresponding low-dose dose-response slope, or “potency,” exhibited by a set of experimental toxicity data. This multi-model approach is first summarized, some concerns about it are discussed in view of constraints on low-dose dose-response information typically available from toxicity studies,

Address correspondence to Dr. K.T. Bogen, Exponent Health Sciences, 500 12th Street, Oakland, CA 94607; kbogen@exponent.com

Generic Hockey-Stick Model

and then its performance when applied to simulated sets of quantal (i.e., dichotomous-response) dose-response data is compared to that of a simpler “Generalized Hockey-Stick” (GHS) model proposed to estimate BMD and potency. Results of these analyses are presented, together with an application of the GHS model, to estimate potential human cancer risk posed by an environmental chemical contaminant known to both increase and reduce cancer incidence in chronically exposed rodents.

1.1 BMDS Modeling Procedure

The BMDS approach is used to characterize toxicity risk for both non-cancer and cancer endpoints based on quantitative analysis of either quantal or continuous dose-response data, and is typically applied to identify or characterize toxicity using data that exhibit a significantly elevated toxicity response and, moreover, represent the most sensitive identified endpoint, species, sex, and strain (EPA 2000a, 2010a b). The resulting characterization is expressed in terms of estimated BMD and its lower (typically one-tail 95%) confidence bound (BMDL), both of which pertain to a user-specified level of “benchmark response” (BMR), in excess of background risk, that is estimated to lie within or near the observed range of response. For many types of quantal toxicity data, BMR = 0.1 typically, by default, and corresponding BMD and BMDL estimates are denoted BMD_{10} and $BMDL_{10}$ (EPA 2000a, 2010b); for brevity, these estimates are denoted herein as d_{10}^* and d_{10}^* , respectively. The dose d_{10}^* , in particular, is used as a “point of departure” (POD) for calculating a corresponding acceptable dose (D^*) (EPA 2000a, 2002, 2010b). For non-cancer endpoints, a combined adjustment/uncertainty factor (UF)—accounting separately for human inter-individual variability, animal-to-human differences, data deficiencies, etc.—is applied to effectively reduce d_{10}^* by defining $D^* = d_{10}^*/UF$ (EPA 2002). For endpoints with a plausible or expected linear low-dose dose-response, acceptable dose $D^* = R^*/q^*$ is typically defined instead in terms of an acceptable level of risk, R^* , and a corresponding upper confidence bound $q^* = BMR/BMDL$ on toxic “potency,” which upper bound is also often referred to as a “slope factor” and denoted SF, in view of the fact that it is the upper bound on the slope of a straight line drawn from the POD to the origin, corrected for background (EPA 2005, pp. 1-14, 3-23). The BMDS procedure has been interpreted and applied widely to characterize potency (slope factors) defined in this way (e.g., Gaylor and Gold 1998; Gaylor 2000; NRC 2000; Gold et al. 2003; Brenner 2004; EPA 2005, 2010b c; Knafla et al. 2006; Simon et al. 2008, 2009; CalEPA DPR 2004; CalEPA OEHHA 2009; Stern 2009). A corresponding implied best estimate of potency in this context is here defined as $q = BMR/BMD$. Over 3,500 users of BMDS are currently registered, approximately 30% of whom are outside the U.S.,

K. T. Bogen

and roughly 25% of whom are in government agencies; among U.S. users, about 20% are from academia (Gift 2009).

1.2 Some Drawbacks of the BMDS Approach

The six-step BMD “decision tree” recommended to evaluate BMDS modeling output does not address or control for procedure-wise error rates that result from its application of multiple statistical procedures. For example, to evaluate output from fits of each of nine primary and/or seven alternative quantal BMDS dose-response models to quantal (dichotomous) data, the software provides multiple options for parameter choices, the decision tree recommends how to interpret the estimates and statistics generated in order to choose among the array of BMDL estimates generated (EPA 2000a, 2010a b). These guidelines address neither how well the recommended modeling procedure actually performs in terms of accuracy or coverage by estimators used in the procedure, nor whether simpler procedures might yield results of similar or greater reliability, nor how the application of multiple statistical tests to multiple model-specific fits affects the statistical interpretation of modeling results obtained. Users are required simply to assume that returned results are meaningful and reliable when applied to relatively small sample sizes like those typical of toxicity data sets.

The BMDS approach relies exclusively on numerical methods to obtain maximum likelihood estimates (MLEs) that have valid and optimal asymptotic properties, but that may be biased when applied to realistic data sets involving relatively few observations. EPA (2000b) recommended “If asymptotic normality cannot be assumed either because the sample size is too small ... or because MLEs were not (or could not) be obtained, bootstrap methods should be employed [... as] a versatile non-parametric method that can be used in a wide variety of situations to obtain the sampling distribution of any model parameter.” Consistent with this recommendation, particularly for small sample sizes (e.g., less than ~30 to 50), parametric bootstrap methods (see DiCiccio and Efron 1996) have been applied as an alternative approach to estimate parameters of quantal dose-response models, (Portier and Hoel 1983; Bailer and Portier 1988; Morris 1988; Smith and Sielkin 1988; Foster and Bischof 1991; Bailer and Smith 1994; Al-Saidy et al. 2003; Nitcheva et al. 2007; Swartout 2007; Zhu et al. 2007). Bootstrap methods remain unavailable through BMDS to estimate model parameter values or confidence limits, as do modified profile likelihood methods that can also reduce bias and improve coverage accuracy for small samples (see, e.g., Barndorff-Nielsen 1983; Brazzale and Davison 2008).

None of models in the BMDS suite typically applied for routine potency or BMD estimation are particularly suited to characterize a hormetic dose-response relationship (i.e., one in which response at least

Generic Hockey-Stick Model

initially declines with increasing dose). Rather, nearly all models assume that risk or detriment in excess of background increases monotonically with increasing dose. BMDS procedures include an analysis of residuals that should identify any clearly anomalous set of data that is inconsistent with this assumption of monotonically increasing risk. Although unconstrained polynomial models offered by BMDS for continuous or quantal data could be applied in any such case, these particular models are not typically considered in regulatory applications, in view of their essentially arbitrary ability to fit a curve to any pattern of data without reference to a plausible mechanistic or biological basis. To the extent there is no intrinsic plausibility of quantal BMDS models applied, the BMDS procedure amounts to an inefficient way to fit an arbitrary smooth curve and confidence bounds through binomially distributed data. To the extent that fits of meaningful BMDS models are affected by data at relatively high dose(s), the modeling effort is effectively “wasted” by addressing data that, by definition, are less relevant or irrelevant to the intended BMD and/or potency (i.e., low-dose dose-response) measure(s) of interest.

1.3 Generic Hockey-Stick (GHS) Model

By definition, only information about response at the lower end of the dose-response curve bears most clearly and directly on BMD and potency. Moreover, it is typically impossible from a statistical standpoint to rule out the possibility that the observed data were sampled from a dose-response relationship that either (1) contains a (non-zero) linear coefficient in dose, or (2) reflects a mixture of two or more functions (e.g., response patterns pertaining to two or more corresponding phenotypes and/or genotypes), one of which contains a (non-zero) linear coefficient in dose. For any risk function $R(d)$ of dose d that is monotonically proportional to a polynomial in d , such as the traditional multistage risk model (Anderson et al. 1983), a linear term $q d$ of that polynomial with $q \neq 0$ must dominate in the limit as $d \rightarrow 0$. Consequently, for data that at some dose(s) exhibit significantly increased risk above background, plausible BMD and potency estimates and their confidence bounds can always be obtained using a suitably generic “hockey-stick” (GHS) version of such a multistage-type risk model. By definition, such a model contains an unconstrained linear coefficient in dose combined with enough additional non-negatively constrained polynomial terms to estimate both BMD and potency aspects of increased risk.

The present study was undertaken to compare BMD and potency estimates obtained using a relatively simple GHS model (described in Methods) to those obtained using the BMDS approach, using quantal dose-response data simulated at specified doses from six specified models of net risk of extra response above corresponding assumed independent

K. T. Bogen

background rates of response. An additional “hormetic” risk function was also considered, simply to illustrate the relative flexibility of a GHS model, and also to characterize negative/hormetic dose-response relationships, and the straightforward approach it offers to test objectively for such possible negativity, either as an attribute of one specific data set, or as a net characteristic implied by a set of related data sets that exhibit a combination of significantly positive and significantly negative dose-response relationships. The latter GHS-model capability was illustrated specifically as described below.

1.4 GHS Estimation of Net Anthraquinone Cancer Risk

Anthraquinone (AQ) increased the incidence of several types of tumor in rats or mice chronically exposed by diet in a National Toxicology Program (NTP) bioassay, but also markedly reduced the incidence of mononuclear cell leukemia (MCL) in male and female rats (NTP 2005). This reduction was considered a direct effect of AQ (NTP 2005; pp. 83, 86, 94):

Several drugs are based on the AQ ring system, including the anthracycline glycosides doxorubicin and daunorubicin, which are used extensively in cancer chemotherapy as well as newer chemotherapeutic agents such as mitoxantrone.... The incidences of MCL were markedly reduced in exposed male and female rats. Although splenic toxicity is often correlated with reduced incidences of MCL ..., it is unlikely that the mild nature of the lesions that occurred in the spleen in the current study could account for the dramatic decrease in incidences. This suggests that the reduction was due to a direct effect of AQ or its metabolite(s) on the development of MCL. Similar decreases have been observed in the 2-year studies of 1-amino-2,4-dibromoanthraquinone and emodin.... Decreased incidences of MCL in male and female rats were attributed to exposure to AQ.

It is noteworthy in this regard that, in 2009, NTP started using Harlan Sprague-Dawley rats for its future studies, due to health-related concerns about their F344N rat colony, including a high incidence of leukemia (King-Herbert and Thayer 2006; King-Herbert et al. 2010). However, EPA (2005) cancer risk assessment methods make no default assumption concerning tumor site-concordance between species; rather, tumor potency exhibited in available bioassay animal data is typically assumed to demonstrate potential cancer potency in humans.

AQ-related compounds can induce apoptotic cell killing by a variety of molecular mechanisms, including inhibition of protein kinase CK2 and/or topoisomerase II (Hartley et al. 1990; Sengupta 1993; De Moliner et al. 2003; Koceva-Chyta et al. 2005), and can directly inhibit human

Generic Hockey-Stick Model

tumor cell proliferation (Cichewicz et al. 2004). It is particularly noteworthy that 1,4-diamino-substituted AQ antitumor agents are cytotoxic to human leukemic cells, and that dose-response kinetics for cell killing examined in detail for 1,5- and 1,8-diamino-substituted anthraquinones in these cells, and in LoVo cells (a human Caucasian colon adenocarcinoma cell line) and Chinese hamster ovary cells exposed to other substituted anthraquinones, clearly exhibit linear-no-threshold-like, “one-hit” kinetics of induced cell killing (Kimler and Cheng 1982; Drewinko et al. 1983; Hartley et al. 1990). Such evidence for linear cell-killing kinetics for AQ-related compounds was the basis for the present illustrative application of a GHS model to characterize observed AQ-induced reductions in rat MCL risk, in view of potential human cancer risk posed by environmental AQ exposures (CalEPA 2006). Specifically, this information was used to combine estimates of AQ-induced tumor induction and suppression in rodents to estimate the potential net carcinogenic potency of AQ to environmentally exposed humans.

2. METHODS

All calculations described were performed using *Mathematica*[®] 7.0.1 software (Wolfram 2010) and related *RiskQ* software (Bogen 2002).

2.1 GHS Model

The GHS model described extends an earlier approach developed to assess and model potential response reduction with increasing dose (Nascarella et al. 2009), by adding polynomial flexibility to a hockey-stick model of quantal-response data that provides quantitative estimates of corresponding BMD and potency and associated estimation errors. Specifically, the linearized multistage model (Anderson et al. 1983) was modified as follows. BMDS software implements a BMD version of the linearized multistage model referred to as the quantal Multistage Cancer (MC) model. Specifically, the GHS function $1 - \exp(-\sum_i q_i d^i)$, $i \in G$ for $G =$ any combination of $\leq g$ elements of $\{0, 1, \dots, g-1, g+1\}$, was used to model the risk or probability $R(d)$ of quantal (i.e., dichotomous) response among n_i individuals in the i^{th} of g dose groups (including the control group) as a function of dose d . Extra risk $A(d)$ over an assumed independent background risk $p_0 = R(0)$ was calculated using Abbot’s correction as $A(d) = [R(d) - p_0] / (1 - p_0)$. In the multistage model, $\text{Max}(i) = g-1$ and all the exponentiated polynomial coefficients q_i are constrained to be non-negative. In contrast, in the GHS model, $\text{Max}(i) = g+1$, and all coefficients q_i are constrained to be non-negative, except for the linear (potency) coefficient, q_1 (for convenience hereafter denoted simply as q), which is constrained only to ensure that $R(d) \geq 0$ for all $d \leq \text{Max}(d_j)$, where $d_j =$ denotes the j^{th} dose ($2 \leq j \leq g$) included in the data set being fit. The GHS model thus can have a slope that is negative as $d \rightarrow 0$, and

K. T. Bogen

otherwise has somewhat greater flexibility than the MC model to reflect more abrupt (albeit, as with the MC model applied using its default assumptions, only monotonically increasing) nonlinearity in dose-response.

Parameters and confidence limits for the GHS model were calculated using a modified version of the method for fitting a multistage model to quantal data described previously (Bogen 1994; Bogen and Witschi 2002), whereby each transformed data set $\{d_j, -\ln(1-R_j)\}$ is fit by completely analytic, nonnegativity-constrained weighted least-squares regressions of all polynomials assuming binomial sampling error, with the best fit defined as that fit (among all good fits obtained by this procedure) which yields the minimum value of chi-square, $\chi^2 = \sum_j (\text{SR}_j)^2 = \sum_j [R_j - n_j R(d_j) + 1/2]^2 / \{n_j R(d_j) [1 - R(d_j)]\}$, with respect to observed data $\{d_j, R_j\}$, where $R(d_j)$ = predicted response at dose d_j and SR_j = the j^{th} standardized residual. The GHS model applies this method to all polynomials implied by sets G defined above, and (as the only GHS procedure that may involve numerical optimization, if necessary) solves one or more convex-polynomial roots to impose the additional constraint mentioned concerning the linear coefficient q . Good fits were defined as those with χ^2 p-values >0.05 and $\text{Max}(|\text{SR}_j|)$ of < 2 , where $\text{Max}(|\text{SR}_j|)$ = the maximum absolute squared residual (MASR). As with the linearized multistage model (Anderson et al. 1983), initially poor fits (if $g > 2$) were re-fit after sequential elimination of the highest dose group until a good fit was obtained. Each fitted GHS model yielded a direct estimate of q , whereupon d_{10} was calculated as the root in dose of $\text{BMR} = 0.1 = A(d_{10})$. Distributions characterizing estimation error in q and d_{10} , from which were calculated the upper one-tail 95% confidence limit (q^*) on potency q and lower one-tail 95% confidence limit (d_{10}^*) on BMD d_{10} , were obtained by the parametric bootstrap method (Bogen 1994; Bogen and Witschi 2002), using 200 sets of data simulated assuming binomial sampling error about the corresponding GHS-predicted response. Best GHS fits to simulated data were each defined as the minimum- χ^2 fit conditional on a χ^2 p-value >0.001 and $\text{MASR} < 3$.

GHS fits could sometimes be improved (e.g., for data sets exhibiting a relatively steep positive or negative initial slope) by imposing the additional constraint that a good fit must include an estimated non-zero linear coefficient, in which case the GHS model was referred to as being “linearized.” Even the linearized GHS model can fail to provide a meaningful estimate of potency if data happen to exhibit a steeply negative dose-response. For quantal data points $\{d_j, R_j\}$ in such cases, an alternative “complementary linearized GHS” procedure instead fits a standard one-hit model $R(d) = 1 - (1-r') \exp(-a d)$ to the complementary data $\{d_j, 1-R_j\}$, and to then estimate the potency b of monotonic risk suppression as $b = a$

Generic Hockey-Stick Model

$r/(1-r)$ pertaining to the original data, where non-complementary background risk r is, in this case, estimated by $r = 1-r'$ (see Appendix).

In the context of the GHS model, potency q (or q^*) is the value (or upper bound) of slope or derivative of $A(d)$ with respect to d in the limit as $d \rightarrow 0$. Conditional on a linear (one-hit) model such as $A(d_{10}) = \text{BMR} = 1 - \exp(-q d_{10})$, and on $q^* > q > 0$, it follows that $d_{10} = -\ln(1-\text{BMR})/q$ and $d_{10}^* = -\ln(1-\text{BMR})/q^*$. Plots comparing the left and right sides of the latter equations were therefore used to assess the extent to which GHS estimates of potency and BMD provide nearly equivalent (hence, essentially redundant) information due to a linear contribution from GHS fits that happen to dominate at doses $\leq d_{10}$.

2.2 Comparison of BMDS vs. GHS Estimates of BMD and Potency

The BMDS approach was implemented using EPA software and corresponding recommended “decision-tree” procedures for interpreting BMDS output for fits to quantal data (EPA 2000a, 2010a b), except that pertaining to visual inspection of plotted fits because this recommended step could not be automated. BMDS methodology was applied (in automated batch-file mode) conditional on model-specific default assumptions, to fit each set of simulated data to the following eight BMDS quantal models: quantal linear (QL), multistage cancer (MC), gamma (GM), logistic (LG) probit (PR), Weibull (WB), log-logistic (logLG), and log-probit (logPR). This subset of the current set of nine dichotomous models offered by BMDS was deemed adequate to characterize BMDS performance for a relatively large set of quantal models typically applied. The BMDS model subset used includes all six of the six non-hormetic risk models used to simulate quantal data (as described below), and excludes the dichotomous Hill model (added to BMDS in 2008) and seven “dichotomous alternative” models offered that differ from corresponding default dichotomous models only insofar as they estimate background in a different way (e.g., background dose instead of background risk). BMDS “decision-tree” steps 1–5 (EPA 2010b) were applied to each data set by: (1) setting $\text{BMR} = 0.1$, (2) fitting each of the eight models listed above (dropping the highest dose group), (3) retaining only “good” fits defined as those with χ^2 goodness-of-fit p-values > 0.1 ; (4) collecting the sorted BMDL and corresponding Akaike Information Criterion (AIC) values (BMDL_j and AIC_j , $j = 1, \dots, k$ with $k \leq 8$) calculated for each j^{th} retained fit; and finally (5) selecting $\text{BMDL}_1 = \text{Min}(\text{BMDL}_j)$ unless (a) BMDL_1 was excluded as an “outlier” or (b) $\text{BMDL}_{\min} \in \text{BMDL}_s$ (for any sorted BMDL subset specified by $s = 1, \dots, m$ with $m \leq k$) with $\text{BMDL}_m / \text{BMDL}_1 \leq 3$, in which case the BMDL associated with $\text{Min}(\text{AIC}_s)$ was selected, or if AIC_s values were all equal, then the geometric mean of all BMDL_s values was selected. In BMDS methodology, Step 5a is addressed, but not defined explicitly (EPA 2010b). To implement Step 5a,

K. T. Bogen

BMDL₁ was defined as an outlier if a p-value of ≤ 0.001 was produced by nested F-test for outliers (Selvin 1995) comparing the variance of the set of log-transformed BMDL_s values with vs. without BMDL₁, provided that $m \geq 3$ and $\text{Max}(AIC_s) - AIC_1 < 2$. That is, if $m < 3$ or the latter AIC difference was ≥ 2 , then BMDL₁ was accepted as a potentially meaningful BMDL estimate. After applying the five steps described, the BMDS quantal model(s) associated with the selected BMDL value was (or were, with equal weight) recorded as the corresponding “BMDL-associated” risk model(s).

BMD and potency estimates obtained by the BMDS and GHS methods described above were compared for 100 sets of data simulated conditional on doses, background (p_0) risks, numbers n_j of exposed animals, and each of the seven quantal risk models summarized in Table 1, assuming binomial sampling error about model-specific predicted risks conditional on n_j . As noted above, six of the seven risk models are quantal models specifically included in BMDS and the seventh is a hormetic quantal response model included to illustrate GHS-model flexibility. These comparisons addressed: difference (bias) between the arithmetic mean of each estimate made by obtaining a (for BMDS estimates, BMDL-associated

TABLE 1. Doses and risk models used to simulate quantal response data.^a

Risk model	Symbol	Doses d_j $j = 1, \dots, 5$ (mg/kg/day)	Background risk, $p_0 = R(0)$ (unitless)	Risk model ^a , $R(d)$, of dose d	True potency ^b , $q/R(d)$ (mg/kg/day) ⁻¹	True BMD ^c , $d_{10}/R(d)$ (mg/kg/day)
Linear	L	0, 1, 2, 4, 10	0.05	$1 - (1-p_0)\exp(-0.04d)$	0.04	2.63
Linear-quadratic	LQ	0, 1, 2, 4, 10	0.05	$1 - (1-p_0)\exp(-0.02d - 0.005d^2)$	0.02	3.01
Probit	PR	0, 1, 2, 4, 10	0.05	$1 - (1-p_0)\Phi[(d-7)/2.5]$	0	3.80
Logistic	LG	0, 1, 2, 4, 10	0.10	$p_0/[p_0 + (1-p_0)\exp(-0.25d)]$	0.0225	2.99
Weibull	WB	0, 1, 2, 4, 10	0.10	$p_0/[p_0 + (1-p_0)\exp(-0.075d^{1.5})]$	0	4.63
Gamma	GM	0, 1, 2, 4, 10	0.10	$1 - (1-p_0)G(1.1, 20, d)$	0	2.74
Hormetic	H	0, 1, 3, 9, 27	0.10	$1 - (1-p_0)\exp(0.04d - 0.004d^2)$	-0.04	12.2

^a A total of $N = 50$ animals was assumed in each dose group. $\exp(x) = e^x$ for any x where e is Euler’s constant; Φ = the cumulative standard normal distribution function; $G(a, b, x)$ = the cumulative distribution function evaluated at x for a generalized gamma distribution with shape parameter a and scale parameter b . A total of $n = 100$ data sets, each i^{th} set (for $i = 1, \dots, n$) containing the five data points $\{Dose_{ij}, Response_{ij}\} = \{d_{ij}, n_{ij}^*/N\}$ (for $j = 1, \dots, 5$), were simulated using each of the indicated models, where each of the simulated n_{ij}^* values was sampled randomly and independently from a binomial distribution with parameters N and $R(d_{ij})$, using risk functions $R(d)$ defined in column 5 each evaluated at corresponding doses $d = d_{ij}$. Models L and LQ have corresponding BMDS quantal dose-response model names: Quantal Linear (QL) and Multistage Cancer (MC), respectively.

^b q = potency (limiting value of increased risk per unit dose as dose approaches zero, assuming an independent background risk); a negative value indicates risk-reducing potency.

^c BMD = BMD₁₀ = d_{10} = benchmark dose assuming a Benchmark Response (BMR) level of 0.10, defined as target excess risk assuming an independent background risk level p_0 , i.e., assuming $BMR = 0.10 = [R(d_{10}) - p_0] / (1 - p_0)$.

Generic Hockey-Stick Model

ed) fit to simulated data in relation to its corresponding expected value listed in Table 1, corresponding standard error of each estimate, the significance of estimated bias as assessed by t-test adjusted (as indicated by a p-value denoted p_{adj}) for multiple (here, seven) independent comparisons using Hommel's modified Bonferroni-type procedure (Wright, 1992), and statistical coverage (a summary measure of performance) by the estimators q^* and d_{10}^* . Also calculated were the percentages of matches between BMDL-associated risk models identified by BMDS and each of the six risk models included in BMDS that were used to generate corresponding model-specific sets of simulated data. Because maximum-likelihood methods are used by BMDS to estimate both d_{10} and d_{10}^* , these percentages must approach 100% for all six of these models asymptotically as $n_i \rightarrow \infty$, and thus divergence from 100% indicates the extent to which asymptotic convergence was not achieved. Similar percentages were also recorded for fits to simulated hormetic data, just to see which BMDS models might tend to be associated (by definition, erroneously) with d_{10}^* estimates obtained by fitting such data.

GSH and BMDS estimates of d_{10} and d_{10}^* obtained for the same sets of data simulated from six risk models (excluding the hormetic model) were compared graphically and also by linear regression, which in the case of d_{10}^* excluded points identified by nested F-test as regression outliers (Selvin 1995).

2.3 GHS Estimation of Net Cancer Potency of Anthraquinone

Absent adequate epidemiological data as reviewed by NTP (2005), net potential human cancer risk of AQ was estimated from NTP bioassay data involving: groups of 50 male and 50 female F344/N rats fed diets containing 0, 469, 938, or 1,875 ppm (for males 0, 20, 45, 90, and 180 mg/kg/day, and for females 0, 25, 50, 100, and 200 mg/kg/day) of AQ for 105 weeks; and groups of 50 male and 50 female B6C3F1 mice fed diets containing 0, 833, 2,500, or 7,500 ppm (for males 0, 90, 265, or 825 mg/kg/day, and for females 0, 80, 235, or 745 mg/kg/day) of AQ for 105 weeks. Survival of all groups of male rats was similar. Survival of exposed groups of female rats was significantly greater than that of corresponding control rats. Mean survival of AQ-exposed female mice equaled or exceeded that of control mice; survival of exposed male mice was reduced significantly only in the highest exposure group; and the highest exposed male mouse dose group also exhibited >3-fold more "natural" deaths than did male control mice (NTP 2005, pp. 8–9). Because survival was not reduced in the rats or female mice studied, and was not reduced in any dosed group of male mice except (significantly) in the highest dose group, time-to-tumor risk models were not considered relevant to characterizing BMD and potency based on the NTP (2005) bioassay data for AQ described. Instead, tumor data were adjusted for intercurrent

K. T. Bogen

mortality differences among dose groups, as is done for the “Poly3” test that is routinely applied by NTP to assess the significance of dose-related effects on tumor incidence (Bailer and Portier 1988). Tumor types analyzed were limited to those tumor sites, or combinations of tumor sites, reported by NTP (2005) to have a tumor incidence that was statistically significantly elevated ($p < 0.05$) in a trend-wise fashion, among those tumor types for which NTP concluded there was biologically meaningful evidence of a dose-related trend. Potencies for benign tumor types were not calculated if the tumor-specific incidence data indicated unambiguously that potency estimated for corresponding combined benign and malignant tumors would substantially dominate that estimated only for benign tumors.

The estimated potency distribution for AQ-reduction of MCL risk was fit using the complementary linearized GHS procedure (Section 2.1) for sex-specific MCL incidence data after adjusting for intercurrent mortality. Resulting estimated distributions of sex-specific potency were given equal weight, and the result was weighted equally with that for the one other significantly (positively) affected rat tumor type. Sex-specific potency distributions for the most elevated tumor type observed in mice were also weighted equally. Resulting species-specific distributions were weighted equally. For each species/sex/type-specific rodent potency q estimated for animals of weight W (kg), a corresponding human-equivalent (HE) potency (q_{HE}) was estimated as $q_{HE} = q [(70 \text{ kg})/W]^{1/3}$ (Anderson et al. 1983; EPA 2005). Net AQ potency (in rodents or humans), aggregating over all tumor-suppressing and all tumor-inducing potencies, was estimated using a modification (see Appendix) of an approach previously described to estimate aggregate excess risk for nonthreshold, quantal, toxic end points caused by exposure to multiple non-hormetic agents, assuming independent actions and background risks (NRC 1994). This modification provides a general approach to calculating the net potency of any set of jointly induced and suppressed toxic risks, assuming independent corresponding background risks.

3. RESULTS

3.1 Comparison of BMDS and GHS Estimates of BMD and Potency

Table 2 summarizes BMD and potency estimation achieved by applying the BMDS procedure to 100 sets of response data at five doses that were simulated based on each of the seven assumed quantal risk models described in Table 1, which include six of the eight quantal BMDS risk models that were fit to each simulated data set, and a hormetic risk model that illustrates a dose-response pattern not typically addressed by any of the quantal BMDS models routinely applied in regulatory-compliance contexts. Thus, a total of 700 data sets were fit to each of eight BMDS

Generic Hockey-Stick Model

TABLE 2. BMDS model output for 100 sets of response data at five doses simulated from different dose-response patterns.

Model used to generate data ^a	n^b	Parameter ^a , Z	Expected value ^a , EZ	\bar{Z}^a	Bias, $\bar{Z}-EZ$	RMCV ^c	Bias-test p-value, P_{adj}^d	\bar{Z}^{*a}	Z^{*e} Coverage ^e
L	97	q	0.04	0.039	-0.0011	0.048	0.57	0.091	0.71
LQ	99	q	0.02	0.035	0.015	0.068	0	0.045	1.
PR	97	q	0	0.030	0.030	0.0007	0	0.038	1.
LG	98	q	0.0225	0.040	0.017	0.083	0	0.063	1.
WB	96	q	0	0.027	0.027	0.0011	0	0.037	1.
GM	97	q	0	0.039	0.039	0.0017	0	0.084	1.
H	65	q	-0.04	0.0082	0.048	0.0057	0	0.011	1. (0 ^e)
L	97	d_{10}	2.63	3.17	0.54	0.059	0.0059	2.10	0.71
LQ	99	d_{10}	3.01	3.19	0.19	0.035	0.15	2.50	0.74
PR	97	d_{10}	3.80	3.58	-0.22	0.098	0.10	2.76	0.94
LG	98	d_{10}	2.99	2.97	-0.02	0.041	0.89	2.21	0.90
WB	96	d_{10}	4.63	4.15	-0.48	0.16	0.020	2.94	0.98
GM	97	d_{10}	2.74	3.06	0.32	0.14	0.084	2.00	0.74
H	65	d_{10}	12.2	12.8	0.66	0.029	0.14	9.17	0.98

^a The seven risk models ($M_{data} = L, LQ, PR, LG, WB, GM, \text{ or } H$) used to generate simulated data, the doses, the parameters (q and d_{10}) and the corresponding model-dependent expected parameter values and corresponding units are defined in Table 1. EZ = the expected value of parameter Z . Symbols \bar{Z} and \bar{Z}^* denote the arithmetic mean values of the BMDS-calculated maximum likelihood estimate of selected-model-specific parameter Z , and of its 1-tail (for q) upper or (for d_{10}) lower 95% confidence limit, respectively, obtained by applying BMDS and its associated "decision tree" steps for model selection to n data sets simulated assuming the indicated risk model.

^b Out of 100 M_{data} -specific sets of simulated data, n = the number of data sets for which the BMDS procedure yielded convergent estimates for parameters q and d_{10} .

^c RMCV = root mean square (RMS) coefficient of variation = $100\%(\text{RMSE}/EZ)$, where RMSE = RMS error estimated as $\text{Sqrt}[\sum\{(Z_i - EZ)^2 / (n_i - 1)\}]$.

^d P_{adj} = p-value from t-test of difference between \bar{Z} and EZ using $n_i - 1$ degrees of freedom, adjusted for multiple (here, seven) such independent data-shape-specific comparisons using Hommel's modified Bonferroni-type procedure (Wright, 1992).

^e Coverage = probability that Z^* is $\geq EZ$ or is $\leq EZ$ for $Z=q$ or $Z=d_{10}$, respectively; value in parentheses = $\text{Prob}(q^* \leq 0)$.

quantal risk models, with convergence on d_{10} and d_{10}^* estimates being successful for nearly all but the hormetic data sets, which exhibited a 35% non-convergence rate. Causes of BMDS convergence failure were not investigated. As expected, estimates of BMDS potency q were quite biased ($p_{adj} = \sim 0$) for data sampled from the hormetic risk model. The same was true for five of the other six risk models used to simulate quantal data, for which corresponding q^* estimates exhibited consistent over-conservatism (coverage = 1, rather than the nominal 0.95 confidence level). The BMDS approach yielded plausibly unbiased ($p_{adj} < 0.01$) estimates of d_{10} and plausibly adequate d_{10}^* coverage (≥ 0.80) for six and four, respectively, of the seven risk models (including the hormetic model) used to simulate quantal data. The BMDS approach yielded d_{10}^* coverage $\geq 90\%$ for only four of the seven risk models used to generate simulated data.

K. T. Bogen

TABLE 3. Percent agreement between the models used to generate simulated data, and BMDL-associated risk models identified by BMDS.

Model (M_{data}) used to generate simulated data ^a	n ^b	BMDL-associated risk model (M_{fit}) (% of fits to n simulated data sets): ^c							
		QL	MC	PR	LG	WB	GM	logPR	logLG
L	97	21.6	4.1	17.5	7.2	3.1	8.2	10.3	27.8
LQ	99	12.5	10.1	38.4	25.3	3.4	4.4	5.1	1.0
PR	97	0	49.1	2.1	28.9	4.1	3.6	9.8	2.1
LG	98	17.2	11.1	26.5	32.7	2.9	3.6	3.1	3.1
WB	96	2.1	19.8	16.7	41.7	0	13.5	3.1	3.1
GM	94	23.3	8.3	10.3	18.6	4.2	3.9	7.2	24.2
H	65	0	47.7	0	3.1	0	46.2	1.5	1.5

^a The seven risk models ($M_{\text{data}} = \text{L, LQ, PR, LG, WB, GM, or H}$) used to generate simulated data, the doses, the parameters (q and d_{10}) and the corresponding model-dependent expected parameter values and corresponding units are defined in Table 1.

^b Out of 100 M_{data} -specific sets of simulated data, n = the number of data sets for which the BMDS procedure yielded convergent estimates for parameters q and d_{10} .

^c BMDS dichotomous dose-response model types (M_{fit}) used to fit to each set of simulated data: QL = Quantal linear, MC = Multistage cancer, PR = Probit, LG = Logistic, WB = Weibull, GM = Gamma, logPR = log-Probit, and logLG = log-Logistic. Percentage values (P_i) listed across all model types in each of the seven M_{data} -specific rows are defined as follows, for $j = 1, \dots, 7$: $P_j = 100\% m_j/n$, where m_j = the number of fits involving the j^{th} M_{data} type. For this calculation, M_{data} values of L and LQ were assumed to be equivalent to M_{fit} values of QL and MS, respectively. Row-specific P_i values may sum to $>100\%$ due to rounding. BMDS fits, and associated P-values, may reflect either multiple M_{fit} types that when optimized had an equivalent form, or a geometric mean of M_{fit} -specific fits that all yielded similar BMDL estimates within a 3-fold factor (see Methods). In each case k total models contributed to a “best fit” M_{fit} to data simulated from the j^{th} M_{data} type, each of the k contributing model types was given a weight of $1/k$ when calculating m_j and P_j . The difference between 100.0 and each P_i value listed in bold typeface indicates the magnitude of non-concordance observed between corresponding values of M_{data} and M_{fit} .

Table 3 summarizes the degree to which BMDL-associated risk models identified by BMDS tended to mis-specify the correct non-hormetic risk model from which simulated data were actually sampled. Rates of correct risk-model specification, shown in the shaded diagonal array of cells in Table 3, were all $<33\%$, with an average value (± 1 SD) of $12\% \pm 13\%$, and was 0% for quantal data simulated using a Weibull (WB) model. Nearly all (65) convergent BMDL-associated quantal models fit to hormetic data were approximately evenly divided between multistage cancer (MC) and gamma (GM) models.

Performance of BMD and potency estimation using the GHS approach is summarized in Table 4. The GHS approach yielded (as defined above) plausibly unbiased estimates of q and plausibly adequate q^* coverage for four and six, respectively, of the seven risk models (including the hormetic model) used to simulate quantal data, and yielded plausibly unbiased estimates of d_{10} and plausibly adequate d_{10}^* coverage for four and seven, respectively, of those seven risk models. The GHS

Generic Hockey-Stick Model

TABLE 4. GHS model output for 100 sets of response data at five doses simulated from different dose-response patterns.

Risk model ^a	Parameter ^a , Z	Expected value ^b ,		Bias, $\bar{Z} - EZ$	RMCV ^a	Bias-test		Z* Coverage ^b
		Z	EZ			p-value, P _{adj} ^a	\bar{Z}^* ^a	
L	q	0.04	0.032	-0.0085	0.055	0.0010	0.053	0.88
LQ	q	0.02	0.027	0.0068	0.16	0.13	0.062	0.95
PR	q	0	0.00051	0.00051	0.0026	0.84	0.036	0.87
LG	q	0.0225	0.025	0.0024	0.14	0.84	0.065	0.90
WB	q	0	-0.0012	-0.0012	0.0031	0.84	0.040	0.73
GM	q	0	0.028	0.028	0.0031	0	0.057	0.98
H	q	-0.04	-0.026	0.014	0.046	0	0.00095	1.00 (0.97)
L	d_{10}	2.63	3.41	0.78	0.063	0.000058	2.05	0.89
LQ	d_{10}	3.01	3.13	0.12	0.045	0.44	1.81	0.91
PR	d_{10}	3.80	3.96	0.16	0.026	0.28	2.61	0.93
LG	d_{10}	2.99	3.37	0.38	0.056	0.099	1.71	0.90
WB	d_{10}	4.63	4.76	0.13	0.037	0.44	2.52	0.96
GM	d_{10}	2.74	3.49	0.74	0.0620	0.00015	1.90	0.94
H	d_{10}	12.2	13.0	0.84	0.015	0.000089	10.3	0.92

^a Risk models (L, LQ, PR, LG, WB, GM, H), doses, the number of animals assumed in each dose group, parameters (q and d_{10}) and corresponding model-dependent expected values of the model parameters (and corresponding parameter units) are all defined in Table 1. Definitions of Z, EZ, RMCV and P_{adj} are given in Table 2. P-values < 10⁻¹⁵ are listed as 0. The GHS model was fit to a total of 700 data sets, consisting of 100 sets simulated assuming each of the seven indicated dose-response function shapes, to obtain the listed corresponding estimates for parameters q and d_{10} . Estimates \bar{Z} and \bar{Z}^* here denote the arithmetic mean values of estimates of parameter Z, and of its 1-tail (for q) upper or (for d_{10}) lower 95% confidence limit, respectively, obtained by fitting the GHS model to n data sets, $S_i = \{d_{ij}, n_{ij}^*\}$ for $i = 1, \dots, n$ and $j = 1, \dots, 5$, each simulated assuming the indicated risk model, with each of the Z_i and corresponding upper/lower bounds Z_i^* calculated as the arithmetic mean and corresponding bound obtained from GHS-fits for 200 sets of data $\{d_{ij}, n_{ij}^{**}\}$ simulated assuming n_{ij}^{**} are distributed binomially with parameters N and n_{ij}^*/N .

^b Coverage = probability that Z^* is $\geq EZ$ or is $\leq EZ$ for $Z=q$ or $Z=d_{10}$, respectively; value shown in parentheses is Prob($q^* \leq 0$).

approach yielded d_{10}^* coverage $\geq 90\%$ for six of the seven risk models used to generate simulated data, and $\geq 89\%$ for all seven models. Although 97% of GHS potency estimates obtained to data sampled from the hormetic risk model were correctly negative, these estimates were also significantly biased ($p_{\text{adj}} < 0.0001$) in a positive (conservative) direction.

Figure 1 compares d_{10} and d_{10}^* estimates from those GHS fits to non-hormetic simulated data yielding positive q or q^* estimates ($n = 283$ or 543, respectively), to functions of the corresponding GSH estimates of q and q^* , respectively, that yield a perfectly linear relationship conditional on a linear (one-hit) risk model, BMR = 0.1. Deviations from linearity among points in the top and bottom plots of Figure 1 indicate the degree to which these GHS fits were substantially nonlinear at the doses in the range of d_{10} and d_{10}^* , respectively, whereas a linear pattern of points indicates the extent to which corresponding BMD- and potency-related estimates provide redundant information. Approximately 91% of the points

K. T. Bogen

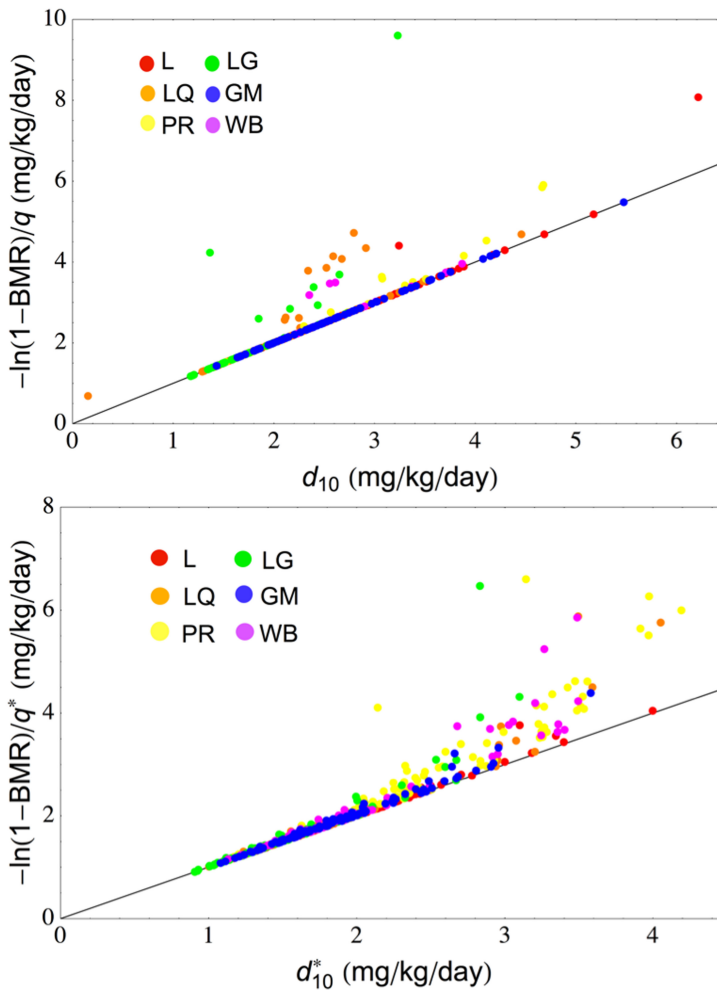


FIGURE 1. GHS estimates of (top) d_{10} and (bottom) d_{10}^* (on X-axis) compared to corresponding values of (top) $-\ln(1-BMR)/q$ and of (bottom) $-\ln(1-BMR)/q^*$ (on Y-axis) predicted conditional on corresponding linear (one-hit) models of increased risk, $BMR = 0.1 = 1 - \exp(-q d_{10})$ and $BMR = 0.1 = 1 - \exp(-q^* d_{10}^*)$, respectively, for those of 100 data sets simulated from each of six different risk models defined in Table 1 that yielded positive estimated values of q ($n = 283$) and q^* ($n = 543$). Black lines show $Y = X$.

shown in each plot of Figure 1 are consistent with a linear pattern to within 15%, and (as expected) all deviations from linearity involve q or q^* estimates that (conditional on a linear response function) over-estimate the corresponding, directly estimated values of d_{10} and d_{10}^* , respectively, generally by a factor of ≤ 1.5 . The q -related deviations from linearity involve primarily fits to data simulated from linear-quadratic (LQ) and logistic (LG) risk models, whereas q^* -related deviations from linearity involve primarily fits to data simulated from probit (PR) and Weibull (WB) risk models.

Generic Hockey-Stick Model

The top and bottom plots of Figure 2 compare d_{10} and d_{10}^* estimates, respectively, obtained using the BMDS approach (on the X-axis) vs. the GHS approach (on the Y-axis) for fits obtained by both methods to non-hormetic simulated data ($n = 584$). The plotted d_{10} and d_{10}^* estimates exhibit highly significant ($p \approx 0$) positive correlations indicated by coefficients of determination (R^2) equal to 0.68 and 0.46, respectively, after excluding as outliers ($F = 2.86$, $df = \{19, 564\}$, $p = 0.000075$) a cluster of 18 BMDS-based d_{10}^* estimates < 0.6 mg/kg/day that were an average (± 1 SD) of 8.4 (5.7) times lower than the corresponding 18 GHS estimates. The d_{10} estimates by both approaches exhibit roughly symmetric scatter with a slope and 95% confidence limits (CL) of 0.98 (0.93, 1.04), consistent with a null hypothesis of estimates by both methods that are, on average, equal. In contrast, for d_{10}^* estimates (even excluding the 18 outlying points), the scatter is asymmetrical with a slope (95% CL) of 0.69 (0.63, 0.75), indicating a significant tendency for d_{10}^* estimates obtained by the GHS approach to be somewhat less (i.e., more conservative) than corresponding BMDS d_{10}^* estimates. More specifically, $\sim 62\%$ of the d_{10}^* points are nearly symmetrically distributed, with a slope (95% CL) of 0.93 (0.88, 0.97) and $R^2 = 0.83$, and the remaining $\sim 35\%$ comprise GHS estimates of d_{10}^* that are each $< 75\%$ of the value of the corresponding BMDS estimate.

3.2 GHS Estimate of Net Anthraquinone Cancer Potency

Tumor types deemed by NTP (2005) to have been affected by chronic dietary exposure of rats and mice to AQ are listed in Table 5, together with corresponding estimates of rodent potency and human-equivalent GHS potency, and associated confidence bounds (q^* and HE q^* , respectively). These tumor types include mononuclear cell leukemia (MCL) in male and female rats; renal cell adenoma or carcinoma (RTAC) in female rats; hepatoblastoma (benign or malignant, HB), hepatocellular carcinoma (HC), and/or hepatocellular adenoma or carcinoma (HAC) in male and/or female mice. The GHS estimates of AQ potency for suppressing MCL in male and female rats are summarized in Figure 3. The estimated MCL potencies for male and female rats are both significantly negative ($p < 0.01$), which illustrates the ability of the GHS model to provide an objective statistical test of whether an agent exhibits a truly negative initial dose-response trend.

The initial GHS fit to all female rat RTAC data yielded a predicted control incidence rate (7.5%) that, assuming binomial sampling error, is statistically inconsistent ($p = 0.00061$) with the corresponding historical control rate of 1/901 (NTP 2005, Table B4a). The listed RTAC potency, obtained after deleting the two highest dose groups, yields a GHS fit predicting 0% incidence in the control group, which is statistically consistent with the historical data. The female mouse control incidence rate of 6/48 (12.5%) of HAC or HB (or HAC alone) was significantly less than the cor-

K. T. Bogen

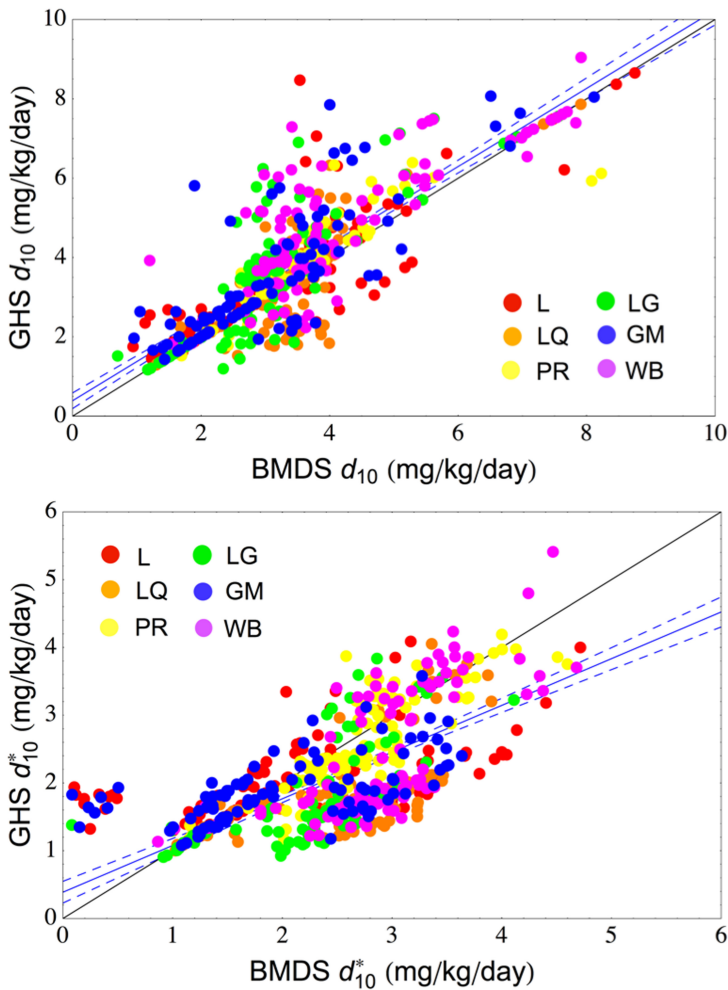


FIGURE 2. Comparison of (top) d_{10} and (bottom) d_{10}^* estimates obtained using the BMDS (on X-axis) vs. the GHS modeling approach (on Y-axis), for ≥ 94 data sets simulated from each of six different risk models defined in Table 1 ($n = 584$). Black lines show $Y = X$. Blue line and dashed curves show each corresponding linear fit and its 95% confidence limits (for d_{10}^* , excluding 18 BMDS estimates all < 0.6 mg/kg/day that are significant outliers).

responding rate of 273/852 (32.0%) exhibited in Battelle Columbus Laboratories' historical control data (NTP 2005, Table D4a) among untreated female B6C3F1 mice ($p = 0.0021$, by 2-tail Fischer exact test). The latter control data exhibited identical incidence rates and range for HAC alone or for combined HAC or HB. The control rate listed for these tumors is also less than the historical control rate for female B6C3F1 mice in all NTP contract laboratories, as reported by NTP (2006) (444/1601 = 27.7% for HAC or HB, 443/1601 = 27.7% for HAC alone, $p = 0.011$ by 2-tail Fischer exact test). Thus, it can be argued that potency information for AQ-induced HAC or HB in female B6C3F1 mice should not be used

Generic Hockey-Stick Model

TABLE 5. GHS-model estimates of AQ tumorigenic potency in rodents.

Species, Sex	Tumor type ^a	q (mg/kg/day) ⁻¹	q^* (mg/kg/day) ⁻¹	q_{HE}^b (mg/kg/day) ⁻¹	q_{HE}^{*b} (mg/kg/day) ⁻¹
Rat, M	MCL	-0.13	(-0.24, -0.050)	-0.68	(-1.3, -0.27)
Rat, F	MCL	-0.052	(-0.11, -0.010)	-0.33	(-0.67, -0.063)
Rat, F	RTAC	0.0034	0.0072	0.021	0.045
Mouse, M	HB	0.00091	0.0018	0.010	0.020
Mouse, M	HC or HB	0.0022	0.0034	0.025	0.038
Mouse, M	HAC or HB	0.0048	0.0090	0.054	0.10
Mouse, F	HC	0.00015	0.00059	0.0016	0.0063
Mouse, F	HAC or HB	0.011	0.017	0.12	0.18

^a MCL = mononuclear cell leukemia, RTAC = renal cell adenoma or carcinoma, HB = hepatoblastoma (benign or malignant), HC = hepatocellular carcinoma, HAC = hepatocellular adenoma or carcinoma.

^b q = tumorigenic potency (limit on increased risk per unit dose d as $d \rightarrow 0$); columns 3 and 5 list estimated expected values of q ; asterisk (*) indicates 1-tail 95% confidence bound(s); HE q -subscript indicates human-equivalent potency, derived assuming $q_{HE} = q(70 \text{ kg}/w)^{1/3}$ where w = adult animal body weight in kg, and w was assumed to be 0.445 and 0.280 kg, and to be 0.050 and 0.055 kg, for male and female rats, and for male and female mice, respectively, used in the NTP (2005) rodent cancer bioassay of AQ.

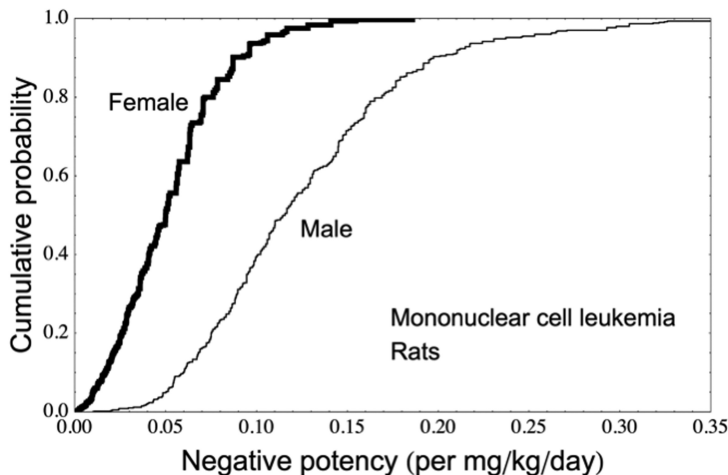


FIGURE 3. Cumulative distribution of anti-tumorigenic AQ potency at suppressing spontaneous mononuclear cell leukemia in male and female F344/N rats, based on GHS-analysis of corresponding NTP (2005) cancer bioassay data.

to estimate potential net potency of AQ in humans, because the corresponding bioassay data are anomalous. However, in the present study, this information was (conservatively) used for GHS-based estimation of net potency, as described in Methods.

The GHS-based estimate of net human-equivalent tumorigenic AQ potency (q_{HE}), shown in Figure 4, has an expected value, and 50th, 95th,

K. T. Bogen

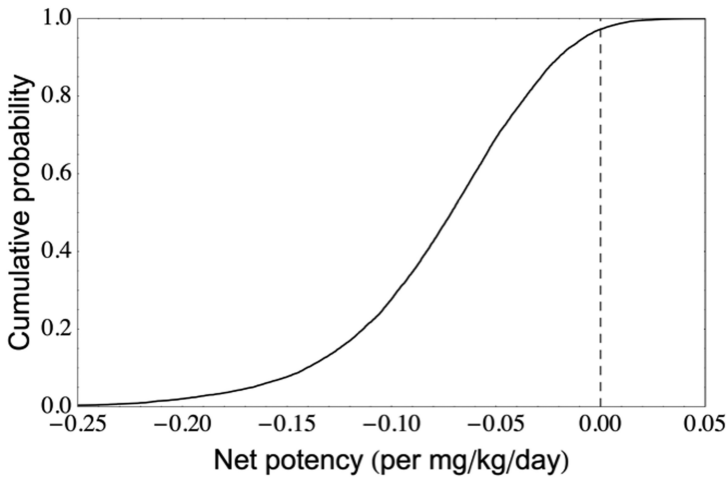


FIGURE 4. GHS-based estimate of net human-equivalent tumorigenic AQ potency (Q_{HE}), calculated as a stochastic difference between (a) the stochastic sum of positive carcinogenic potencies estimated from data indicating the ability of AQ to induce renal tubular cell adenomas or carcinomas in female F344/N rats and hepatoblastomas in male B6C3F1 mice, and (b) estimated AQ potency at suppressing spontaneous leukemia in F344/N rats, after first applying interspecies adjustments to all estimated rodent potencies involved. Dashed vertical line corresponds to $Q_{HE} = 0$.

and 97.5th values of -0.077 , -0.071 , -0.0080 and 0.0018 ($\text{mg}/\text{kg}/\text{day}$)⁻¹, respectively, and indicates that $\text{Prob}(q_{HE} > 0) = 0.028$. The latter probability illustrates how the GHS model can be combined with the method described in the Appendix to assess the likelihood of net positive or net negative potency of an agent or exposure scenario that jointly induces and suppresses cancer risk.

4. DISCUSSION

The GHS model is simpler to use to estimate BMD and potency aspects of low-dose dose-response than the multi-model BMDS numerical optimization and decision-tree procedure (EPA 2000a, 2010b), for five key reasons:

- 1. Number of Models.** The GHS model consists of one dose-response model applied in one automated step to each data set, whereas BMDS consists of a suite of many models.
- 2. Focus on Low-Dose Dose-Response.** By virtue of its mathematical form, the GHS model focuses more efficiently on the specific problem of characterizing available low-dose dose-response information of key relevance to making decisions concerning protection from or prevention of toxicity, without expending additional modeling effort, or model-output analysis effort, to accurately characterize dose-response over the entire range of possible response. The GHS approach is thus a conceptually simpler approach that achieves the same purpose as

Generic Hockey-Stick Model

BMDS, albeit by focusing more narrowly on the low-dose end of available dose-response information. Dose-response characterization over an entire response range can be quite important in a number of decision contexts (e.g., involving experimental design, pharmacology, resource prioritization, or triage) that hinge on accurate prediction over this entire range (see, e.g., Bogen 2005). The BMDS approach can effectively support such goals, whereas the GHS approach described cannot.

- 3. Efficiency of Analysis.** BMDS risk models must be fit individually to each data set. Non-automated decision-tree methods must then be applied to the complete set of model-specific BMDS outputs, to address model-specific and inter-model goodness-of-fit criteria and select a desired model output from a complex array of model outputs produced for each model fit. While the relatively new BMDS “sessions template” feature automates the application of multiple models within certain predefined BMDS-model subsets, it does not automate either the application of the decision-tree criteria or the recommended additional step by which users are asked to document how the criteria were applied to sets of BMDS output. Nor does the current version of BMDS automatically re-fit poorly fitting models to data sets that delete the highest dose group(s). The “sessions template” feature also requires a considerable amount of labor to set up required model-specific input-data and model-execution files, all requiring interaction with a mildly labor-intensive Excel[®]-like user interface, which has very few symbolic or object-oriented features that might enhance implementation efficiency.
- 4. Mathematical Implementation.** The GHS model described was optimized nearly entirely analytically, using a numerical procedure (if necessary) only to solve for roots of convex polynomials of degree >4 , whereas BMDS is optimized by standard numerical methods to maximize likelihood functions or “surfaces” with respect to the parameters of each model. For small samples, nonlinear models, and/or models with constrained parameters, such likelihood surfaces may be multimodal and/or otherwise depart substantially from quadratic forms ideally suited for numerical optimization. Consequently, the BMDS approach may occasionally (or, evidently for hormetic data, fairly often) fail to converge or may converge on meaningless estimates. In such cases, users are advised to restart optimization using different initial values of parameters to be estimated. In contrast, the GHS model always converges, to yield parameter estimates. Of course, standard maximum likelihood methods, or newer modified/adjusted profile likelihood methods, could also be used to optimize the GHS model, in which case convergence failure would arise for this approach as well.

K. T. Bogen

5. Characterization of Potential Hormetic Trend (i.e., Negative Potency).

The GHS model is structured to perform an automatic statistical characterization of the likelihood of any hormetic trend (i.e., negative slope) exhibited by low-dose dose-response data, as a routine component of performing a GHS model fit. Except for an unconstrained polynomial model that is not routinely applied due to its mechanistic implausibility, BMDS models do not characterize or assess the likelihood of hormetic behavior the way these models are typically applied. The “complementary” procedure applied to obtain GSH fits to data on AQ-induced reduction of MCL in rats could also be applied using the BMDS approach. However, the complementary procedure was required in the AQ illustration only because the hormetic responses were so pronounced as to drive tumor risks to near zero in all dosed groups of rats. Had the reduction been a bit less pronounced, the GHS model could reflect this successfully, whereas BMDS could not do as well consistently. This is because simply releasing the constraint on the linear term of, for example, the BMDS MC or polynomial model might do well to model some hormetic data sets, but could fail to give meaningful results for others by predicting negative risks within the observed dose range, because, unlike the GHS model, BMDS models do not (currently) impose functional constraints to ensure that $R(d) \geq 0$. Methods discussed in the Appendix extend the ability of the GHS model to characterize net potency associated with multiple independent endpoints.

GHS-based estimates of BMD and potency obtained for simulated non-hormetic data sets were found to be largely redundant, in the sense that these two measures of low-dose dose-response conveyed essentially equivalent information (Figure 1). This result raises a fundamental question concerning the entire rationale of using a BMD approach. If BMD and BMDL can nearly always be predicted very accurately, or otherwise typically at least fairly accurately (within a factor of 1.5), from corresponding estimates of potency and upper-bound potency, respectively, and if substantial deviations from this predictive relationship are in each case statistically supported by available low-dose data, why not base decisions solely on the estimated value or upper bound of the maximum slope (i.e., potency) of low-dose dose-response that is statistically consistent with those data? A key point of the GHS model is that by applying it, relevant available data can be used to estimate potency or upper-bound potency in a way that is statistically consistent with those data. By applying the GHS model, these estimates are not pre-judged by inferring any limitation on their meaning or reliability by reference to a BMD or BMDL that is in turn defined by a BMR selected without reference to the data being fit. A focus on estimating slope *per se*, which is provided by the GHS approach, does not *necessarily* imply a belief, or policy of inferring, that a low-dose

Generic Hockey-Stick Model

slope estimate that is statistically consistent with a set of dose-response data is meaningfully interpreted at doses far below the dose associated with the lowest observed response above background. Such a belief or policy is typically adopted for ionizing radiation, radiomimetic chemical carcinogens, and many or most other genotoxic chemical carcinogens (e.g., EPA 2005). For other (presumed “nonlinear”) endpoints, the predictive potency/BMD relationship observed in this study could be applied to implement the usual method of deriving a reference dose/concentration (EPA 2002), and to help formulate unified risk-based approaches to regulating environmental exposures associated with different types of toxic endpoints (NRC 2009).

GHS-based estimates of BMD and potency performed about as well or better than corresponding BMDS estimates, at least for quantal data of the type investigated (Tables 2 and 4). Specifically, the GHS approach clearly outperformed the BMDS approach in estimating potency and its upper confidence limit, which is understandable insofar as the BMDS approach was not designed to estimate this low-dose dose-response characteristic, despite widespread use of the BMDS approach to do just this (see Introduction). The BMDS approach yielded plausibly unbiased estimates of BMD for more (six of seven) types of simulated quantal data than did the GHS approach (four of seven types). However, the GHS approach yielded fairly good BMDL coverage (≥ 0.89) for all seven data types simulated, while the BMDS approach did so for only four of the seven data types. Risk acceptability decisions based on BMD methods typically are based on potency or reference dose/concentration calculations involving BMDL rather than BMD estimates (e.g., Cal EPA DPR 2004; EPA 2000a, 2002, 2005; CalEPA OEHHA 2009).

BMD estimates and approximately 62% of BMDL estimates obtained by both methods applied to simulated sets of non-hormetic data were highly correlated (Figure 2). Approximately 35% of BMDL estimates obtained by the GHS approach were <75% below (i.e., more conservative than) corresponding BMDS estimates, and approximately 3% of BMDLs estimated by BMDS had very low values far below the corresponding GHS estimates (Figure 2). The BMDS approach failed, on average, about 90% of the time to correctly identify BMDL-associated quantal dose-response models that were used to simulate corresponding sets of non-hormetic dose-response data sets analyzed, which all were generated assuming 50 animals in each of five dose groups (Tables 1 and 3). The sample size used in this study exceeds that of many if not most toxicity data sets used for regulatory risk assessment. The very low rate of successful model identification indicates that even this sample size is far from approximating asymptotic conditions under which maximum-likelihood BMDS methods guarantee a 100% success rate. The low success rate does not indicate any fundamental flaw in the BMDS approach, because it was not specifically

K. T. Bogen

optimized to correctly match models used to estimate BMDL with those that actually generate data to be fit. Rather, the low success rate indicates the degree to which model-specific information turned out not to be relevant to estimates obtained by using the BMDS approach.

Modifications could be made to the BMDS-recommended model-evaluation decision tree (EPA 2000a, 2010b) to enhance its reliability and performance. Such modifications were not investigated in view of the adequate performance of the proposed, simpler GHS approach to modeling quantal data, which can readily be extended to address the case of continuous data.

The application of the GHS model to anthraquinone data, together with the net-potency-calculation method described in the Appendix, illustrate how these approaches can be used to help assess whether a net positive or negative cancer risk is posed by environmental or dietary exposures to one or more agents that, singly or jointly, exhibit an ability to both induce and suppress cancer risks. Agents or agent mixtures with such a capacity currently pose a regulatory dilemma, insofar as no consensus exists on how (or whether) to perform quantitative net-potency assessment for such agents, rather than simply ignoring demonstrated tumor-suppressing capacity. From a public health perspective, risk management must necessarily consider indirect (unintended) imposition or augmentation of net expected harm, if and whenever it may be likely to occur—a point emphasized in a key National Research Council recommendation (NRC 1994). The key question upon which a determination of significant net cancer risk hinges is whether or not hypothesized significant anti-carcinogenic effects are plausibly induced at low environmental exposure levels with a linear no-threshold dose-response; that is, with the same type of low-dose dose-response relationship that is typically assumed for many, if not most, chemical carcinogens. This appears to be likely in the case of AQ-induced MCL suppression (see Introduction). For this type of agent, such information must be incorporated into net-potency estimation in order to implement the National Research Council recommendation that quantitative probabilistic approaches be used to ensure that regulatory decisions inflict no public health detriment (NRC 1994).

APPENDIX:

Net Potency of Jointly Induced and Suppressed Toxicity

The NRC (1994) previously described a method for calculating aggregate potency for increasing the risk of nonthreshold, quantal, toxic end points caused by exposure to multiple agents—or a single agent causing multiple related endpoints such as multiple tumor types—assuming independent actions and background risks, conditional on a multistage

Generic Hockey-Stick Model

model of exposure-related risk (NRC 1994). This general approach can be modified as follows to allow similar calculations in the case that potencies characterize both induction and suppression of the endpoints involved, when quantal dose-response data for each i^{th} (among n_i) induced endpoint A_i are well described for inducing agents by multistage-type (including GHS) risk models, $R_i(d) = 1 - \exp(-\sum_{k=0} q_{i,k} d^k)$, and for each j^{th} (among n_j) monotonically suppressed endpoints B_j by corresponding complementary risk models $R_j(d) = \exp(-\sum_{k=0} q_{j,k} d^k)$ that in the simplest case are conditioned on $k \leq 1$ and so reflect risks that decline exponentially. Let q_i and a_j denote the linear coefficients in dose of these models applied to a specified agent or mixture. Note that risk-inducing potencies q_i may each or all have some likelihood of being negative (e.g., if they are estimated using a GHS model). In contrast, it is assumed here that $\text{Prob}(a_j < 0) = 0$ for all j . At sufficiently low dose, these risk models become virtually equivalent to one-hit risk models for risk induction (R) and for monotonic risk suppression (S),

$$R_i(d) = p_i + (1-p_i) [1 - \exp(-q_i d)] \text{ for endpoints } A_i,$$

and

$$S_j(d) = r_j \exp(-a_j d) \text{ for endpoints } B_j,$$

respectively, where p_i and r_j are corresponding background risks conditional on $d = 0$. Since by de Morgan's rule (where each overbar denotes the corresponding complementary set):

$$\begin{aligned} \text{Prob}(A_i \cup B_j) &= 1 - \text{Prob}(\overline{A_i} \cap \overline{B_j}) \\ &= \prod_{i=1}^{n_i} (1 - R_i(d)) \prod_{j=1}^{n_j} (1 - S_j(d)) \\ &= 1 - \prod_{i=1}^{n_i} ((1 - p_i) \exp(-q_i d)) \prod_{j=1}^{n_j} (1 - r_j \exp(-a_j d)), \end{aligned}$$

and because increased risk assuming independent background is $A(d) = [P(d) - P(0)] / [1 - P(0)]$ for any risk model $P(d)$, it follows that increased aggregate risk $A_{A \cup B}(d)$ for endpoints A_i or B_j is:

$$A_{A \cup B}(d) = 1 - \exp\left(-d \sum_{i=1}^{n_i} q_i\right) \times \prod_{j=1}^{n_j} \left(1 - \frac{r_j \exp(-a_j d)}{1 - r_j}\right)$$

K. T. Bogen

Net aggregate potency Q , defined as the derivative (slope) of $A_{\text{AUB}}(d)$ with respect to dose d in the limit as $d \rightarrow 0$, is thus:

$$Q = \sum_{i=1}^{n_i} q_i - \sum_{j=1}^{n_j} \frac{r_j}{1-r_j} a_j = \sum_{i=1}^{n_i} q_i - \sum_{j=1}^{n_j} b_j,$$

where terms b_j here denote effective potencies of monotonic risk suppression estimated for endpoints B_j , $j = 1, \dots, n_j$, respectively. In this expression for aggregate potency Q , whereas the contributing risk-inducing potencies q_i are all independent of corresponding estimated background risks p_i , this is not true of the contributing effective potencies b_j for monotonic risk suppression, which are non-linearly proportional to corresponding background risks r_j . The distribution characterizing uncertainty in Q associated with estimation errors in corresponding parameters q_i and b_j is obtained by evaluating the right-hand expression probabilistically (e.g., by Monte Carlo simulation using random samples from respective distributions estimated for those parameters).

ACKNOWLEDGEMENTS

Many thanks to Paul Booth at Exponent, who prepared DOS batch files that enabled automated applications of EPA BMDS software with multiple sets of input data and model specifications used in this study, and for reviewer comments that improved this manuscript.

REFERENCES

- Al-Saidy OM, Piegorsch WW, West RW, Nitcheva DK. 2003. Confidence bands for low-dose risk estimation with quantal response data. *Biometrics* 59(4):1056–1062.
- Anderson EL, Albert RE, McGaughy R, Anderson L, Bayard S, Bayliss D, Chert C, Chu M, Gibb H, Haberman B, Hiremath C, Singh D, Thorslund T. 1983. Quantitative approaches in use to assess cancer risk. *Risk Anal* 3:277–295.
- Bailer J, Portier C. 1988. Effect of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* 44:417–431.
- Bailer AJ, Smith RJ. 1994. Estimating upper confidence limits for extra risk in quantal multistage models. *Risk Anal* 14(6):1001–1010.
- Barndorff-Nielsen O. 1983. On a formula for the distribution of the maximum likelihood estimator. *Biometrika* 70(2):343–365.
- Bogen KT. 1994. Cancer potencies of heterocyclic amines found in cooked foods. *Food Chem Toxicol* 32:505_515.
- Bogen KT. 2002. *RiskQ 4.2: An Interactive Approach to Probability, Uncertainty and Statistics for use with Mathematica®*. UCRL-MA-110232 Rev. 3. Lawrence Livermore National Laboratory, Livermore, CA.
- Bogen KT. 2005. Risk analysis for environmental health triage. *Risk Anal*; 25:1085–1095.
- Bogen KT, Witschi HP. 2002. Lung tumors in A/J mice exposed to environmental tobacco smoke: Estimated potency and implied human risk. *Carcinogenesis* 23:511–519.
- Brazzale AR, Davison AC. 2008. Accurate parametric inference for small samples. *Statist Sci* 23(4):465–484.
- Brenner TC. 2004. Brief survey of EPA standard-setting and health assessment. *Environ Sci Technol* 38(13):3457–3464.

Generic Hockey-Stick Model

- California Environmental Protection Agency (CalEPA), Department of Pesticide Registration (DPR). 2004. Guidance for Benchmark Dose (BMD) Approach - Continuous Data. DPR MT-2, September 2004. CalEPA DPR, Health Assessment Section, Medical Toxicology Branch, Sacramento, CA. <http://www.cdpr.ca.gov/docs/risk/bmdcont.pdf>
- California Environmental Protection Agency (CalEPA), Office of Environmental Health Hazard Assessment (OEHHA). 2006. Chemical Meeting the Criteria for Listing as Causing Cancer Via the Authoritative Bodies Mechanism: Anthraquinone. Package 19a.5, July 21, 2006. CalEPA OEHHA, Oakland, CA.
- California Environmental Protection Agency (CalEPA), Office of Environmental Health Hazard Assessment (OEHHA). 2009. Technical Support Document for Cancer Potency Factors: Methodologies for Derivation, Listing of Available Values, and Adjustments to Allow for Early Life Stage Exposures. May 2009.
- Cichewicz RH, Zhang Y, Seeram NP, Nair MG. 2004. Inhibition of human tumor cell proliferation by novel anthraquinones from daylilies. *Life Sci* 74(14):1791–1799.
- De Moliner E, Moro S, Sarno S, Zagotto G, Zanotti G, Pinna LA, Battistutta R. 2003. Inhibition of protein kinase CK2 by anthraquinone-related compounds. A structural insight. *J Biol Chem* 278(3):1831–1836.
- DiCiccio TJ, Efron B. 1996. Bootstrap confidence intervals. *Statist Sci* 11(3):189–228.
- Drewinko B, Yang L-Y, Barlogi B, Trujillo JH. 1983. Comparative cytotoxicity of bisantrene, mitoxantrone, ametantrone, dihydroxyanthracenedione, dihydroxyanthracenedione diacetate, and doxorubicin on human cells in vitro. *Cancer Res* 43:2648–2653.
- Foster DH, Bischof WF. 1991. Thresholds from psychometric functions: superiority of bootstrap to incremental and probit variance estimators. *Psychol Bull* 109(1):152–159.
- Gaylor DW, Gold LS. 1998. Regulatory cancer risk assessment based on a quick estimate of a benchmark dose derived from the maximum tolerated dose. *Regulatory Toxicol Pharmacol* 28(3):222–225.
- Gaylor DW. 2000. New issues in carcinogen risk assessment. *Drug Metab Rev* 32(2):187–192.
- Gift J. 2009. U.S. Environmental Protection Agency, National Center for Environmental Assessment, Research Triangle Park, NC (personal communication, 2 Nov).
- Gold LS, Gaylor DW, Slone TH. 2003. Comparison of cancer risk estimates based on a variety of risk assessment methodologies. *Regulatory Toxicol Pharmacol* 37(1):45–53.
- Hartley JA, Forrow SM, Souami RL, Reszka K, Lown JW. 1990. Photosensitization of human leukemic cells by anthracenedione antitumor agents. *Cancer Res* 50:1936–1940.
- Knafla A, Phillips KA, Brecher RW, Petrovic S, Richardson M. 2006. Development of a dermal slope factor for benzo[*a*]pyrene. *Regulatory Toxicol Pharmacol* 45(2):159–168.
- Kimler BF, Cheng CC. 1982. Comparison of the effects of dihydroxyanthraquinone and adriamycin on the survival of cultured Chinese hamster cells. *Cancer Res* 42:3631–3636.
- King-Herbert AP, Thayer. 2006. NTP workshop: Animal models for the NTP rodent cancer bioassay: stocks and strains—Should we switch? *Toxicol Pathol* 34(6):802–805.
- King-Herbert AP, Sills RC, Bucher JR. 2010. Commentary: Update on animal models for NTP studies. *Toxicol Pathol* 38(1):180–181.
- Koceva-Chyta A, Jedrzejczak M, Skierski J, Kania K, Jozwiak Z. 2005. Mechanisms of induction of apoptosis by anthraquinone anticancer drugs aclarubicin and mitoxantrone in comparison with doxorubicin: Relation to drug cytotoxicity and caspase-3 activation. *Apoptosis* 10(6):1497–1514.
- Morris MD. 1988. Small-sample confidence limits for parameters under inequality constraints with application to quantal bioassay. *Biometrics* 44(4):1083–1092.
- Nascarella MA, Stanek EJ, Hoffmann GR, Calabrese EJ. 2009. Quantification of hormesis in anti-cancer-agent dose-responses. *Dose-Response* 7:160–171.
- National Research Council (NRC). 1983. *Risk Assessment in the Federal Government: Managing the Process*. National Academy Press, Washington, DC.
- National Research Council (NRC). 1994. *Science and Judgment in Risk Assessment*. Appendix I-1: Aggregate Risk of Nonthreshold, Quantal, Toxic End Points Caused by Exposure to Multiple Agents (Assuming Independent Actions). National Academies Press, Washington, DC, pp. 516–518.
- National Research Council (NRC). 2000. *Methods for Developing Spacecraft Water Exposure Guidelines*. National Academies Press, Washington, DC.
- National Research Council (NRC). 2009. *Risk and Decision Making*. National Academies Press, Washington, DC.

K. T. Bogen

- National Toxicology Program (NTP). 2005. Toxicology and Carcinogenesis Studies of Anthraquinone (CAS No. 84-65-1) in F344/N Rats and B6C3F1 Mice (feed studies). NTP TR 494, NIH Publication No. 05-3953. U.S. Department of Health and Human Services, National Institutes of Health, Research Triangle Park, NC.
- National Toxicology Program (NTP). 2006. NTP Toxicology Database Management System. NTP Historical Controls Report, All Routes and Vehicles: Mice. U.S. Department of Health and Human Services, National Institutes of Health, Research Triangle Park, NC, pp. 46–47.
- Nitcheva D, Piegorsch W, West RW. 2007. On use of the multistage dose-response model for assessing laboratory animal carcinogenicity. *Regul Toxicol Pharmacol* 48(2):135–147.
- Portier C, Hoel D. 1983. Low-dose-rate extrapolation using the multistage model. *Biometrics* 39:897–906.
- Sengupta SK. 1993. Inhibitors of DNA-transcribing enzymes. In: Foye WE (ed), *Cancer Chemotherapeutic Agents*. American Chemical Society, Washington, DC, pp 205-260.
- Selvin S. 1995. *Practical Biostatistical Methods*. Duxbury Press, Wadsworth Publishing Co., Belmont, CA, pp. 18-20, 71-72.
- Simon T, Kirman CR, Aylward LL, Budinsky RA, Rowlands JC, Long TF. 2008. Estimates of cancer potency of 2,3,4,7,8-pentachlorodibenzofuran using both nonlinear and linear approaches. *Toxicol Sci* 106(2):519–537.
- Simon T, Aylward LL, Kirman CR, Rowlands JC, Budinsky RA. 2009. Estimates of cancer potency of 2,3,7,8-tetrachlorodibenzo(p)dioxin using linear and nonlinear dose-response modeling and toxicokinetics. *Toxicol Sci* 112(2):490–506.
- Smith LA, Sielkin RL Jr. 1988. Bootstrap bounds for “safe” doses in the multistage cancer dose-response model. *Commun Statist Simul* 17:153–175.
- Stern AH. 2009. Derivation of an ingestion-based soil remediation criterion for Cr⁶⁺ based on the NTP chronic bioassay data for sodium dichromate dihydrate. New Jersey Office of Science, Research Project Summary, June 2009. <http://www.state.nj.us/dep/dsr/chromium/ingestion-cr.pdf> (accessed 17 Aug 2010).
- Swartout J. 2007. Analysis of dose-response uncertainty using benchmark dose modeling. Paper presented at Resources for the Future by Jeff Swartout, U.S. Environmental Protection Agency, National Center for Environmental Assessment, Washington, DC. <http://www.rff.org/Events/Documents/0710%20Swartout.pdf> (accessed 17 Aug 2010).
- U.S. Environmental Protection Agency (EPA). 2000a. Benchmark Dose Technical Guidance Document. EPA/630/R-00/0001F. EPA Risk Assessment Forum, Washington, DC.
- U.S. Environmental Protection Agency (EPA). 2000b. Options for Development of Parametric Probability Distributions for Exposure Factors. EPA/600/R-00/058. EPA National Center for Environmental Assessment and Office of Research and Development, Washington, DC, p. 1–22.
- U.S. Environmental Protection Agency (EPA). 2002. A Review of the Reference Dose and Reference Concentration Processes. EPA/630/P-02/002F, December 2002. EPA National Center for Environmental Assessment and Office of Research and Development, Washington, DC.
- U.S. Environmental Protection Agency (EPA). 2005. Guidelines for Carcinogen Risk Assessment. Federal Register 70(66):177650–18717. <http://www.epa.gov/cancerguidelines> (accessed Aug 1, 2009).
- U.S. Environmental Protection Agency (EPA). 2010a. Benchmark Dose Software (BMDS) Version 2.1.2 (Build 60). (11 June 2010). U.S. EPA National Center for Environmental Assessment, Washington, DC. <http://www.epa.gov/ncea/bmds/dwnldu.html> (last updated on August 13, 2010).
- U.S. Environmental Protection Agency (EPA). 2010b. Benchmark Dose (BMD) Methodology. U.S. EPA National Center for Environmental Assessment, Washington, DC. http://www.epa.gov/ncea/bmds/bmds_training/methodology/intro.htm (last updated on June 16, 2010).
- U.S. Environmental Protection Agency (EPA). 2010c. Toxicological Review of Acrylamide (CAS No. 76-06-1) in Support of Summary Information on the Integrated Risk Information System (IRIS), March 2010. EPA/635/R-07/009F. U.S. EPA, Washington, DC.
- Wright SP. 1992. Adjusted p-values for simultaneous inference. *Biometrics* 48:1005–1013.
- Wolfram Research. 2010. Wolfram *Mathematica*® 7 Documentation Center. Wolfram Research, Inc., Champaign, IL (www.wolfram.com), <http://reference.wolfram.com/mathematica/guide/Mathematica.html> (accessed 26 January 2010).
- Zhu Y, Wang T, Jelsovsky JZ. 2007. Bootstrap estimation of benchmark doses and confidence limits with clustered quantal data. *Risk Anal* 27(2):447–465.