Published in final edited form as:

Stat Methodol. 2008 May ; 5(3): 209–219. doi:10.1016/j.stamet.2007.08.001.

Simultaneous Confidence Bands for Abbott-Adjusted Quantal Response Models

Brooke E. Buckley^a and Walter W. Piegorsch^b

^a Department of Mathematics, Northern Kentucky University, Highland Heights, KY 41099, USA

^b Interdisciplinary Program in Statistics, BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA

Abstract

We study use of a Scheffé-style simultaneous confidence band as applied to low-dose risk estimation with quantal response data. We consider two formulations for the dose-response risk function, an Abbott-adjusted Weibull model and an Abbott-adjusted log-logistic model. Using the simultaneous construction, we derive methods for estimating upper confidence limits on predicted extra risk and, by inverting the upper bands on risk, lower bounds on the benchmark dose, or BMD, at which a specific level of ‘benchmark risk’ is attained. Monte Carlo evaluations explore the operating characteristics of the simultaneous limits.

Keywords

Benchmark analysis; Benchmark dose; Carcinogenic risk assessment; Weibull dose-response model; Log-logistic dose-response model; Quantitative risk analysis

1. Introduction and Background

An important concern in quantitative risk assessment is the characterization and estimation of adverse effects after exposures to hazardous stimuli. The field of risk analysis employs statistical methodology to assess the risks of such exposures, where a primary objective is quantitative characterization of the severity and likelihood of damage to humans or to the environment caused by the hazardous agents [9].

To support quantitative operations in this area, we report on methods for assessing risks at low doses of a toxic stimulus. In keeping with standard terminology, we define risk, $R(d)$, as the probability of some pre-defined adverse effect, such as death, weight loss, birth defect, cancer, or mutation exhibited in a subject exposed to a particular dose level, d , of a toxic agent [5, 19]. We also assume that $R(d)$ is a monotone increasing function of dose. Although seemingly straightforward, this definition contains an important, implicit feature: non-zero risk may exist, even change, for very small levels of d . This extends earlier concepts of risk where, at least for many non-cancer endpoints, one assumes that some dose threshold exists below which $R(d) = 0$. By modeling the risk more formally, however, a richer variety of possible dose-response functions and consequent statistical machinery becomes available [22].

Correspondence to: Brooke E. Buckley.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Notice that $R(0)$ represents the risk to all subjects in a population, independent of exposure status. To construct inferences that adjust for this spontaneous response rate, excess risk functions are often employed for risk estimation. Common with quantal response data is the *extra risk*, $R_E(d)$, defined as the risk above background among subjects who would not have responded under control conditions: $R_E(d) = \{R(d) - R(0)\} / \{1 - R(0)\}$ [33]. An important quantitative objective in risk analysis is proper estimation of $R_E(d)$ and from this, construction of (upper) confidence limits on the extra risk at one or more values of $d \geq 0$. Inversion of this relationship is also possible: if a risk assessor has in mind one or more “benchmark risk” (BMR) levels above which the extra risk becomes unacceptable, one can ask what level of dose leads to attainment of those BMRs. Within this context, the *Benchmark Dose* (BMD) is the smallest value of d that yields a specified BMR [10]. A $100(1 - \alpha)\%$ lower confidence limit on the BMD is called the *Benchmark Dose (Lower) Limit*, or BMDL [11]. It is common in modern risk assessment to select a specific BMDL as a *point of departure* which is then reduced by a set of uncertainty/safety factors to arrive at an acceptable level of exposure or to otherwise establish low-exposure guidelines [17].

We denote the estimated benchmark dose under the posited dose-response model as $\hat{BMD}$. To identify at which specific risk benchmark, BMR, the BMD is determined, we use the notation BMD_{100BMR} , $\hat{BMD}_{100BMR}$, $BMDL_{100BMR}$, etc. Often, more than one BMDL is considered for such a point of departure; e.g., BMDLs corresponding to $BMR = 0.01, 0.05$ and 0.10 might all be calculated and considered for a given adverse effect [16,20,23,29]. This can occur when a risk regulator has a variety of values of BMR in mind for a given data set — each BMR corresponding to a different regulatory or scientific concern — or when a series of BMR values is realized to be of regulatory interest after the data are observed. In these cases, proper statistical practice obliges the user to apply some form of multiplicity correction in the calculation of the confidence limits (including adjustments for the case when decisions are made in a *post hoc* fashion), although to our knowledge this is not common. To overcome the multiplicity/*post hoc* problem, Al-Saidy *et al.* [3] proposed methodology that allows a risk assessor to calculate a simultaneous (upper) $100(1 - \alpha)\%$ confidence *band* on an extra risk function and, using this band, to find simultaneous lower $100(1 - \alpha)\%$ confidence limits on any number of BMDs. That is, risk regulators can make *a priori* or *a posteriori* decisions on any number of benchmark responses for the agent under study and obtain simultaneous lower bounds on the exposure levels producing these benchmark responses. Throughout the effort, valid $100(1 - \alpha)\%$ inferences are guaranteed by the method’s underlying simultaneity. Without such simultaneous protection, consideration of several candidate BMDL values will result in lower-than-stated or otherwise invalid levels of confidence in the dose values eventually selected.

In their study, Al-Saidy *et al.* [3] focused on one specific form of dose-response model, the multistage model from carcinogenicity testing: $R(d) = 1 - \exp\{-\beta_0 - \beta_1 d - \dots - \beta_k d^k\}$, where $\beta_j \geq 0$ for all j . Since this form represents a well-known and popular model for cancer dose-response modeling, their study was a useful first step in developing simultaneous inferences for BMD analysis. A larger issue remained open, however: how and to what degree could their results be extended to other forms of dose-response model(s) useful in quantitative risk assessment? We address these issues by developing simultaneous confidence bands for low-dose risk estimation with a selection of competitors to the multistage model, based on a large class of dose-response forms known as *Abbott-adjusted models*. Section 2 discusses this class and addresses estimation and inferences for R_E . Section 3 introduces the Scheffé-style band, while also addressing BMD/BMDLs. Section 4 studies small-sample features of the methods via Monte Carlo evaluations, and includes illustrations of the approach. Section 5 concludes with a brief discussion.

2. Models and Risk Estimation

In risk-analytic dose-response studies the observations are often in the form of proportions, also called “quantal” data, where subjects are classified in a binary fashion as exhibiting or not exhibiting the adverse health effect. This is particularly common under the cancer risk assessment paradigm highlighted by Al-Saidy *et al.* [3]. When the data appear as proportions, say Y_i/N_i , it is common to assume that the parent distribution of each numerator Y_i is binomial, with sample-size parameter N_i (number of subjects tested). We denote the unknown probability that an individual subject will respond as π_i , so that $Y_i \sim \text{indep. binomial}(N_i, \pi_i)$; i indexes the dose level, $i = 1, \dots, n$. The response probability is a natural quantity for describing the risk, and hence is used here as the risk function, i.e., $R(d_i) = \pi_i$.

To model a dose-response, R is linked to π_i via some monotone function of d . By way of background, it is common in risk assessment to model the probability of an adverse effect (risk) as some core predictor function in dose [4,8,19,21]. Of particular interest is the linear predictor $\beta_1 + \beta_2 x$ where x is a function of d : $x = x(d)$. Then, some nonlinear function extends the core predictor into a probability, $R(d)$. Generalizing this, we model $R(d)$ in the following form [12]

$$R(d) = \beta_0 + (1 - \beta_0)F(d|\beta), \quad (1)$$

where $0 \leq \beta_0 < 1$ represents the background risk, β is the parameter vector associated with the linear predictor, and $F(d|\beta)$ is any monotone function of $d \geq 0$ such that $F(0|\beta) = 0$ and $F(d|\beta) \leq 1$. The presence of non-zero background risk in (1) is an application of Abbott’s formula wherein the action of the hazard at dose d is assumed independent of any background effect [1], hence $R(d)$ is referred to here as an *Abbott-adjusted model*.

In (1), it is common to use some cumulative distribution function for $F(d|\beta)$, as many of these functions satisfy the necessary conditions placed here on $F(\cdot|\cdot)$. With this comprehensive form, the extra risk becomes:

$$R_E(d) = \frac{\beta_0 + (1 - \beta_0)F(d|\beta) - \beta_0}{1 - \beta_0} \quad (2)$$

which clearly simplifies to $F(d|\beta)$.

For purposes of low-dose estimation, we will focus on two forms of Abbott-adjusted risk, each based upon common cumulative distribution functions. The first form replaces $F(d|\beta)$ with the Weibull c.d.f. yielding the following probability of response:

$$R(d) = \beta_0 + (1 - \beta_0) \left(1 - \exp \left\{ -e^{\beta_1} d^{\beta_2} \right\} \right) \quad (3)$$

where $0 \leq \beta_0 < 1$ and $\beta_2 \geq 1$. This model is quite popular in cancer risk assessment and, expanding on the work of Al-Saidy *et al.* [3], is a natural form to use within the context of our benchmark risk application. (The Weibull form is also popular in many non-cancer risk applications, however; see, e.g., [15] or [32].)

A frequently-seen competitor to the Weibull-based model employs a logistic c.d.f. for $F(d|\beta)$ in (1). To make it comparable to the Weibull model in (3), one often sees it on a (natural) log-dose scale, leading to an Abbott-adjusted log-logistic regression model:

$$R(d) = \beta_0 + (1 - \beta_0) (1 + \exp \{-\beta_1 - \beta_2 \ln(d)\})^{-1} \quad (4)$$

where again $0 \leq \beta_0 < 1$ and $\beta_2 \geq 1$. Other expressions for $F(d|\beta)$ in (1) are also possible, of course, although to keep the focus manageable for our risk analytic setting we will restrict attention to these two, popular forms.

Under our models the MLEs, $\mathbf{b} = [b_0 \ b_1 \ b_2]^t$, of the unknown parameters, $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2]^t$, are found by constrained optimization. The operations can be programmed in the software package **R** using its *constrOptim* function [27], in SAS via PROC NLMIXED [34], or in the U.S. Environmental Protection Agency's standalone BMDS software [31]. Krewski and Van Ryzin [22] note that the MLEs have an asymptotic normal distribution when certain, standard regularity conditions are met. From these, appeal to ML invariance [7, Sec. 7.2] produces MLEs for the extra risk under (3) as $\hat{R}_E(d) = 1 - \exp\{-\exp\{-b_1 - b_2 \ln(d)\}\}$, and under (4) as $\hat{R}_E(d) = (1 + \exp\{-b_1 - b_2 \ln(d)\})^{-1}$.

Given the presence of the linear predictor $\beta_1 + \beta_2 \ln(d)$ in our formulations for the extra risk we can appeal to the results of Al-Saidy et al. [3], where simultaneous upper confidence bands for $R_E(d)$ are built over some continuous interval $0 < d < B$. (The value of $B > 0$ is specified *a priori*.) The approach employs a modification of Scheffé's S-method [28] for building confidence bands, and hence we refer to it as a Scheffé-style band. The upper Scheffé-style band on $\beta_1 + \beta_2 \ln(d)$ is given by:

$$b_1 + b_2 \ln(d) + k_\alpha \sqrt{v_{11} + 2v_{12} \ln(d) + v_{22} \ln(d)^2}, \quad (5)$$

where $v_{11} = \text{Var}(b_1)$, $v_{22} = \text{Var}(b_2)$, and $v_{12} = \text{Cov}(b_1, b_2)$ are the large-sample variances and covariance of the MLEs. These quantities are estimated from the inverse of the expected information. The critical point, k_α , is calculated via an algorithm described by Pan *et al.* [25]. The critical point calculation operates on dose values between any two (possibly infinite) limits, A and B . Note that for use in either the Weibull or log-logistic setting, our band must employ a critical point over the range $-\infty < \ln(d) < \ln(B)$. This is a valid application of the Pan *et al.* algorithm, and so may be applied here.

Scheffé-style confidence bands on $R_E(d)$ for the Weibull and log-logistic models are found after appropriate manipulation of (5). (The details are straightforward.) In the case of the Weibull model, this produces the following upper band on $R_E(d)$:

$$R_E(d) < 1 - \exp\left\{-\exp\left\{b_1 + b_2 \ln(d) + k_\alpha \sqrt{v_{11} + 2v_{12} \ln(d) + v_{22} \ln(d)^2}\right\}\right\} \quad (6)$$

Similar manipulations lead to a Scheffé-style upper band on $R_E(d)$ for the log-logistic model:

$$R_E(d) < \left[1 + \exp\left\{-b_1 - b_2 \ln(d) - k_\alpha \sqrt{v_{11} + 2v_{12} \ln(d) + v_{22} \ln(d)^2}\right\}\right]^{-1} \quad (7)$$

In (6) and (7), we appeal to the asymptotic features of the MLEs to validate use of the Pan *et al.* algorithm. The small-sample coverage characteristics of these Scheffé-style bands will be discussed below in Section 4.

3. Benchmark Dose Estimation

Under the Weibull model in (3), the BMD is that value of d that solves $R_E(d) = \text{BMR}$, i.e., $1 - \exp\{-\exp[-\beta_1 - \beta_2 \ln(d)]\} = \text{BMR}$ at a given benchmark risk, $\text{BMR} \in (0, 1)$. Clearly, solving for d gives

$$\text{BMD}_{100\text{BMR}} = \exp\left[\frac{\ln(-\ln(1 - \text{BMR})) - \beta_1}{\beta_2}\right]. \quad (8)$$

If risk is modeled as in (4), then the benchmark dose is the value of d that solves $R_E(d) = 1/(1 + \exp\{-\beta_1 - \beta_2 \ln(d)\}) = \text{BMR}$. This yields

$$BMD_{100BMR} = \exp \left[\frac{\ln(BMR/1 - BMR) - \beta_1}{\beta_2} \right]. \quad (9)$$

Via ML invariance, the MLE, $\hat{BMD}_{100BMR}$, is found by substituting the MLE b_j for β_j wherever it appears in either (8) or (9).

With a point estimate of the benchmark dose, we now seek lower simultaneous confidence limits, denoted above as BMDLs. By inverting the simultaneous upper confidence band on $R_E(d)$, simultaneous lower confidence limits on the BMD for achieving a specified level of risk, BMR, are obtained.

For example, in the Weibull setting we set a specified BMR equal to the upper band in (6) and solve for d . This leads to

$$\ln[-\ln[1 - BMR]] = b_1 + b_2 \ln(d) + k_\alpha \sqrt{v_{11} + 2v_{12} \ln(d) + v_{22} \ln(d)^2} \quad (10)$$

The smallest positive root of (10) is a simultaneous $BMDL_{100BMR}$ under the Abbott-adjusted Weibull model. Similarly, if we equate a specific BMR with the upper band given in (7) for the log-logistic model, we find

$$\ln \left[\frac{BMR}{1 - BMR} \right] = b_1 + b_2 \ln(d) + k_\alpha \sqrt{v_{11} + 2v_{12} \ln(d) + v_{22} \ln(d)^2} \quad (11)$$

The smallest positive root of (11) is a simultaneous $BMDL_{100BMR}$ under our Abbott-adjusted log-logistic model. Solving (10) or (11) for $\ln(d)$ and exponentiating each side guarantees at least one positive root if and only if $4k_\alpha^2[k_\alpha^2(v_{12}^2 - v_{22}v_{11}) + b_2^2v_{11} + 2c_0b_2v_{12} + c_0^2v_{22}] \geq 0$, where

$$c_0 = \ln\{-\ln(1 - BMR) - b_1\}$$

under the Weibull formulation, and

$$c_0 = \ln \left[\frac{BMR}{1 - BMR} \right] - b_1$$

under the log-logistic formulation. Although these conditions are often met in practice, it is unfortunately not the case that existence of the BMDL is guaranteed under either model.

4. Small-Sample Usage

4.1 Monte Carlo Evaluations

Nominal $1 - \alpha$ coverage for our Scheffé-style upper band is based on appeal to asymptotic approximations. Hence in large samples we expect the simultaneous limits to contain the true value of $R_E(d)$ or the true BMD approximately $100(1 - \alpha)\%$ of the time. In small samples, however, the coverage may be less certain. To evaluate this, we undertook a Monte Carlo study of the small-sample simultaneous coverage associated with our Scheffé-style upper bands over a variety of Abbott-adjusted Weibull and log-logistic quantal response models. Specifications for β in each model were chosen to mimic those used in a similar small-sample study of simultaneous BMDLs by Nitcheva, et al. [24], where the background risk, $R(0)$, in these models ranged between 1% and 30%, and the corresponding risk at the highest dose had rates between 10% and 90%. We added one additional parameterization in order to capture a background risk of $R(0) = 0.05$ and a maximum response of 50%. Our six parameterizations are summarized in Table 1.

Four dose levels, $d = 0, 0.25, 0.5, 1$, with equal numbers of subjects, $N_i = N$, per dose-group were used in the simulations, corresponding to a common design in cancer risk experimentation [26]. We selected values of N that ranged over $N = 25, 50, 100, 300, 500, 1000, 5000$, and $10,000$. Of course, in cancer risk assessment, study designs with only the first few of these sample sizes would typically be employed (although, in certain microbiological studies with cells in culture, sample sizes of 1000 and even 5000 would not be unheard of). The larger sample sizes were included here primarily to validate the asymptotic features of the large-sample constructions.

Under each model configuration 2000 pseudo-binomial data sets were simulated, and for each set simultaneous $100(1 - \alpha)\%$ upper confidence bands were calculated on $R_E(d)$ over $0 < d < B$ using the methodology of Section 2. (Technically, the lower limit for the Pan *et al.* algorithm is $\ln(0) \rightarrow -\infty$, but for calculational purposes we set this to -1×10^{10} . The Pan *et al.* upper limit was set to the natural log of the highest dose level: $\ln(1) = 0$.) We report empirical coverage results at $\alpha = 0.05$. Notice then that the approximate standard error of the estimated coverage near $\alpha = 0.05$ is $\sqrt{(0.05)(0.95)/2000} \approx 0.005$, and it never exceeds $\sqrt{(0.5)(0.5)/2000} = 0.011$.

Empirical coverage probabilities for the Scheffé-style upper band when applied to the Abbott-adjusted Weibull models are reported in Table 2; results for the Abbott-adjusted log-logistic models appear in Table 3. From the tables, we see that the Scheffé-style band produces very conservative results. For smaller sample sizes, the coverage is essentially 100% for all models studied, with Parameterization F the exception. In looking at the models, one notes that Parameterization F has the most pronounced increase in probability over the dose range. This appears to provide some stability to the method being studied. For the other parameterizations, a sample size of between 1,000 and 5,000 at each dose was often required to achieve coverage probabilities close to nominal. Similar results were found by Nitcheva *et al.* [24] when applying the Scheffé-style upper band to the multistage cancer dose-response model mentioned in Section 1.

4.2. Illustrative example with BMDLs: Aflatoxin

Aflatoxin B₁ is a common mycotoxin contaminant in corn, peanuts, and cottonseed. In a study of the toxin's potential mammalian carcinogenicity, Wogan, et al. [35] report rates of hepatocellular tumors (as proportions of animals exhibiting the tumor) in male rats after exposure to six different doses of the toxin (in $\mu\text{g/kg/day}$). The doses used in the study were $d_1 = 0.0$, $d_2 = 0.04$, $d_3 = 0.20$, $d_4 = 0.60$, $d_5 = 2.0$, and $d_6 = 4.0$. The corresponding observed proportions were 0/18, 2/22, 1/22, 4/21, 20/25, and 28/28. Cancer risk analysts might wish to build BMDs (and BMDLs) from these data to use as points of departure in a formal risk assessment. Here, we illustrate how our Scheffé-style band can be employed to facilitate this effort.

Applying both the Weibull and log-logistic models to the data, we achieved roughly similar qualitative results: a plot of predicted values versus the original proportions (not shown) gives a reasonably good fit for either model, although this is not altogether surprising when attempting to fit three parameters on six response points. To provide greater discrimination between the two model fits, we appealed to Akaike's [2] Information Criterion (AIC) in its lower-is-better form. We found the AIC for the Weibull model to be $\text{AIC}_W = 75.712$, while for the log-logistic model it was $\text{AIC}_L = 77.253$. [We also applied the multistage cancer dose-response model mentioned in Section 1 to these data, following up on the analysis by Al-Saidy *et al.* [3]. The AIC there was $\text{AIC}_M = 75.729$.] Thus the Abbott-adjusted Weibull form produced a slightly better AIC and we chose to continue our analysis under that model.

Under the Weibull form in (3), the constrained MLEs are $b_0 = 0.0451$, $b_1 = -0.9067$, and $b_2 = 1.9978$. Using the typical BMR suite of 0.01, 0.05, and 0.10, we find the associated point

estimates to be $\widehat{\text{BMD}}_{01} = 0.15743 \mu\text{g/kg/day}$, $\widehat{\text{BMD}}_{05} = 0.35597 \mu\text{g/kg/day}$, and $\widehat{\text{BMD}}_{10} = 0.51039 \mu\text{g/kg/day}$. Simultaneous 95% BMDLs corresponding to these values can then be determined using the Scheffé-style approach presented in Section 3. For the aflatoxin data, we find $v_{11} = 0.1717$, $v_{12} = -0.1810$, $v_{22} = 0.2781$, and $k_{0.05} = 1.843$, producing simultaneous 95% limits of $\text{BMDL}_{01} = 0.01447 \mu\text{g/kg/day}$, $\text{BMDL}_{05} = 0.07026 \mu\text{g/kg/day}$, and $\text{BMDL}_{10} = 0.14069 \mu\text{g/kg/day}$. These values all hold with simultaneous 95% coverage, allowing a risk/safety assessor to apply them in a simultaneous fashion when developing exposure assessments for this toxic agent.

4.3. Illustrative example with Extra Risk: Carbendazim

Despite our focus on cancer as an important risk analytic endpoint, other adverse outcomes receive significant attention in contemporary toxicological risk assessment. Much of this interest centers currently on genetic endpoints, where technological advances have allowed for high throughput and analysis of large numbers of cells, chromosomes, gene markers, etc. As a result, per-group sample sizes can reach into the thousands. For example, Bently et al. [6] report quantal-response data on per-cell chromosomal damage (as non-disjunction events) in human lymphocytes after *in vitro* exposure of the cells to carbendazim, a potentially potent metabolite of the agricultural fungicide benomyl. Since it is relatively easy to acquire and analyze large numbers of healthy lymphocytes, per-dose sample sizes can grow quickly in such studies; Bently et al. operated with $N = 2000$ cells per dose level. Thirteen doses (in ng/ml) were used: $d_1 = 0$, $d_2 = 300$, $d_3 = 400$, $d_4 = 500$, $d_5 = 600$, $d_6 = 700$, $d_7 = 800$, $d_8 = 900$, $d_9 = 1000$, $d_{10} = 1100$, $d_{11} = 1200$, $d_{12} = 1400$, and $d_{13} = 1500$. The resulting quantal response proportions representing damage to chromosome number 8 were, respectively, 16/2000, 20/2000, 24/2000, 9/2000, 19/2000, 34/2000, 44/2000, 54/2000, 82/2000, 77/2000, 54/2000, 108/2000, and 110/2000.

Of interest is estimation of the extra risk of lymphocyte damage as exposure dose increases. Applying both the Weibull and log-logistic models to these data, we found that both models fit the data evenly well: the AIC for the Weibull model was $\text{AIC}_W = 5834.831$, while for the log-logistic model it was $\text{AIC}_L = 5834.664$. Since the log-logistic AIC was slightly better, we continued the fit with an Abbott-adjusted log-logistic model; this produced constrained MLEs of $b_0 = 0.0069$, $b_1 = -18.8473$, and $b_2 = 2.1802$. The corresponding MLE for the extra risk function was thus $\hat{R}_E(d) = (1 + \exp\{18.8473 - 2.1802 \ln(d)\})^{-1}$, with pertinent variance-covariance values given as $v_{11} = 3.9908$, $v_{12} = -0.5530$, and $v_{22} = 0.0767$.

For inferential purposes, we can build a simultaneous 95% confidence band on R_E based on (7). With such large sample sizes, and based on our simulation results from §4.1, we are fairly confident that the simultaneous coverage level here is near to this nominal value. Constrained to the dose range $0 < d < 1500$ (so that the upper limit on d is simply the highest dose from the study), this yields a Scheffé-style critical point of $k_{0.05} = 2.1220$ and subsequent Scheffé-style upper band of

$$R_E(d) < \left[1 + \exp\{18.8473 - 2.1802 \ln(d) - 2.1220 \sqrt{3.9908 - 1.106 \ln(d) + 0.0767 \ln(d)^2}\} \right]^{-1}.$$

Figure 1 plots the two functions. We see that while clearly significant, the extra risk function is very small in magnitude over this dose range, approaching only about 6% at the higher doses. Thus the extra risk of genetic damage appears to be fairly low over the dose range studied with this chemical stimulus.

5. Discussion

Herein, we consider a general class of models, known as Abbott-adjusted models, for use in quantitative risk analysis [13]. Placing emphasis on the cancer risk assessment setting, we study two specific parameterizations of this general class, a Weibull-based and a log-logistic-based dose-response model. Inferences are centered on a Scheffé-style simultaneous upper band for extra risk functions, $R_E(d)$, which are then manipulated to also provide simultaneous lower limits on the benchmark dose (BMD) that leads to a specified benchmark (extra) risk (BMR). Based on a short Monte Carlo study, we find that all the methods for constructing limits on R_E or on BMD exhibit conservative simultaneous coverage for small sample sizes, but that the simultaneous coverage drops towards nominal levels as the sample sizes increase. Risk analysts can apply our results to construct simultaneous inferences on a wider choice of models for the observed quantal response, expanding beyond the multistage form seen in Al-Saidy et al. [3]. This extended model choice can lead to improved risk analytic decision-making in toxicological and other adverse-event risk assessments.

The method's simultaneity is perhaps most useful in the common instance where a variety of BMRs are considered. The simultaneous nature of the underlying bands allows for the different BMDLs calculated at each BMR to still possess minimal $100(1 - \alpha)\%$ confidence for any number of desired BMDLs, even if selected in a *post hoc* manner. Hence, even in the case where the number of multiple benchmark risks/BMDLs is fixed in advance — and thus a simpler and potentially-more accurate Bonferroni-type correction for the multiple inferences might be applied — the added opportunity to (i) make simultaneous inferences on the R_E function and (ii) allow for the possibility of *post hoc* risk levels to be examined, is made available through use of the simultaneous confidence bands. Indeed, Nitcheva et al. [24] found that for the multistage dose-response model mentioned in Section 1, Scheffé-style bands often competed favorably against use of Bonferroni corrections for BMDL acquisition: although the Bonferroni adjustment generally produced tighter bounds, the Scheffé bands were not substantially smaller (more conservative) in many practical instances.

Of course, some caveats and qualifications are in order. The entire notion of benchmark analysis and BMD estimation is model dependent, and other models for $R(d)$, and in particular other choices for $F(d | \beta)$ in (1), may produce different inferences on the BMD. Still, studies have shown that the benchmark method can confer several significant advantages for risk/safety assessment [14], and is generally more flexible than the older, model-independent approach of using so-called No-Adverse-Observed-Effect Levels (NOAELs) for low-dose estimation; see, e.g., Crump [10] or Gaylor and Kodell [18], among many others. It is also important to warn the user that unless very low response rates are observed at a broad spectrum of doses close to zero, it is difficult to apply these benchmark methods at very small BMRs. Indeed, the U.S. EPA in its Technical Guidance Document for BMD estimation [30] suggests that target BMRs should lie between 0.01 and 0.10, and we generally concur with this recommendation.

Acknowledgements

Portions of this research were conducted while both authors were with the University of South Carolina Department of Statistics. Thanks are due to Drs. R. Webster West, Brian T. Habing, Obaid M. Al-Saidy, Daniela K. Nitcheva, Ralph L. Kodell, and an anonymous reviewer for their helpful suggestions. This work was funded under grant #R01-CA76031 from the U.S. National Cancer Institute, grant #RD-83241901 from the U.S. Environmental Protection Agency, and as part of the research arm of the U.S. Department of Homeland Security's Center of Excellence for the Study of Terrorism and Responses to Terrorism (START). Its contents are solely the responsibility of the authors and do not necessarily reflect the official views of these various agencies.

References

1. Abbott WS. A method of computing the effectiveness of an insecticide. *Journal of Economic Entomology* 1925;18:265–267.
2. Akaike, H. Information theory and an extension of the maximum likelihood principle. In: Petrov, BN.; Csaki, B., editors. *Proceedings of the Second International Symposium on Information Theory*. Akademiai Kiado; Budapest: 1973. p. 267–281.
3. Al-Saidy OM, Piegorsch WW, West RW, Nitcheva DK. Confidence bands for low-dose risk estimation with quantal response data. *Biometrics* 2003;59:1056–1062. [PubMed: 14969485]
4. Bailer AJ, Oris JT. Modeling reproductive toxicity in *Ceriodaphnia* tests. *Environmental Toxicology and Chemistry* 1993;12:787–791.
5. Bailer AJ, See K. Individual-based risk estimation for count responses. *Environmental Toxicology and Chemistry* 1998;17:530–533.
6. Bently KS, Kirkland D, Murphy M, Marshall R. Evaluation of thresholds for benomyl- and carbendazim-induced aneuploidy in cultured human lymphocytes using fluorescence in situ hybridization. *Mutation Research* 2000;464:41–51. [PubMed: 10633176]
7. Casella, G.; Berger, RL. *Statistical Inference*. Duxbury; Pacific Grove, CA: 2002.
8. Chen JJ, Gaylor DW. Dose-response modeling of quantitative response for risk assessment. *Communications in Statistics — Theory and Methods* 1992;21:2367–2381.
9. Coherssen, JJ.; Covello, VT. Executive Office of the President; Washington, DC: 1989. *Risk Analysis: A Guide to Principles and Methods for Analyzing Health and Environmental Risks*.
10. Crump KS. A new method for determining allowable daily intake. *Fundamental and Applied Toxicology* 1984;4:854–871. [PubMed: 6510615]
11. Crump KS. Calculation of benchmark doses from continuous data. *Risk Analysis* 1995;15:79–89.
12. Edler L, Kopp-Schneider A. Statistical models for low dose exposure. *Mutation Research* 1998;405:227–236. [PubMed: 9748596]
13. Edler, L.; Kopp-Schneider, A.; Heinzl, H. Dose-response modeling. In: Edler, L.; Kitsos, CP., editors. *Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment*. John Wiley & Sons; Chichester: 2005. p. 5–23.
14. Faustman EM, Bartell SM. Review of noncancer risk assessment: Applications of benchmark dose methods. *Human and Ecological Risk Assessment* 1997;3:893–920.
15. Foronda NM, Fowles J, Smith N, Taylor M, Temple W. A benchmark dose analysis for sodium monofluoroacetate (1080) using dichotomous toxicity data. *Regulatory Toxicology and Pharmacology* 2007;47:84–89. [PubMed: 16965845]
16. Gaylor DW, Aylward LL. An evaluation of benchmark dose methodology for non-cancer continuous-data health effects in animals due to exposures to dioxin (TCDD). *Regulatory Toxicology and Pharmacology* 2004;40:9–17. [PubMed: 15265602]
17. Gaylor DW, Kodell RL. Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. *Risk Analysis* 2000;20:245–250. [PubMed: 10859783]
18. Gaylor DW, Kodell RL. A procedure for developing risk-based reference doses. *Regulatory Toxicology and Pharmacology* 2002;35:137–141. [PubMed: 12051999]
19. Gaylor DW, Slikker WL. Risk assessment for neurotoxic effects. *NeuroToxicology* 1990;11:211–218. [PubMed: 1700346]
20. Karita K, Yano E, Dakeishi M, Iwata T, Murata K. Benchmark dose of lead inducing anemia at the workplace. *Risk Analysis* 2005;25:957–962. [PubMed: 16268943]
21. Kodell RL, West RW. Upper confidence intervals on excess risk for quantitative responses. *Risk Analysis* 1993;13:177–182. [PubMed: 8502790]
22. Krewski, D.; van Ryzin, J. Dose response models for quantal response toxicity data. In: Csörgö, M.; Dawson, DA.; Rao, JNK.; Saleh, AKME., editors. *Statistics and Related Topics*. North-Holland, Amsterdam: 1981. p. 201–231.
23. Markowski VP, Zareba G, Stern S, Cox C, Weiss B. Altered operant responding for motor reinforcement and the determination of benchmark doses following perinatal exposure to low-level

- 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Environmental Health Perspectives* 2001;109:621–627. [PubMed: 11445517]
24. Nitcheva DK, Piegorsch WW, West RW, Kodell RL. Multiplicity-adjusted inferences in risk assessment: Benchmark analysis with quantal response data. *Biometrics* 2005;61:277–286. [PubMed: 15737104]
 25. Pan W, Piegorsch WW, West RW. Exact one-sided simultaneous confidence bands via Uusipaikka's method. *Annals of the Institute of Statistical Mathematics* 2003;55:243–250.
 26. Portier CJ. Biostatistical issues in the design and analysis of animal carcinogenicity experiments. *Environmental Health Perspectives* 1994;102(Suppl 1):5–8. [PubMed: 8187725]
 27. R Development Core Team, R. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2005.
 28. Scheffé H. A method for judging all contrasts in the analysis of variance. *Biometrika* 1953;40:87–104.
 29. Schlosser PM, Lilly PD, Conolly RB, Janszen DB, Kimbell JS. Benchmark dose risk assessment for formaldehyde using airflow modeling and a single-compartment, DNA-protein cross-link dosimetry model to estimate human equivalent doses. *Risk Analysis* 2003;23:473–487. [PubMed: 12836840]
 30. U.S. EPA. Benchmark Dose Technical Guidance Document. U.S. Environmental Protection Agency; Washington, DC: 2000.
 31. U.S. EPA. Help Manual for Benchmark Dose Software Version 1.3, National Center for Environmental Assessment. U.S. Environmental Protection Agency; Research Triangle Park, NC: 2001.
 32. van Wijngaarden E, Beck C, Shamlaye CF, Cernichiari E, Davidson PW, Myers GJ, Clarkson TW. Benchmark concentrations for methyl mercury obtained from the 9-year follow-up of the Seychelles Child Development Study. *NeuroToxicology* 2006;27:702–709. [PubMed: 16806480]
 33. West, RW.; Piegorsch, WW. Extra risk. In: El-Shaarawi, AH.; Piegorsch, WW., editors. *Encyclopedia of Environmetrics*. 2. John Wiley & Sons; Chichester: 2002. p. 731-732.
 34. Wheeler, MW. Benchmark dose estimation using SAS[®]. In: Nelson, GS., editor. *Proceedings of the Thirtieth Annual SAS[®] Users Group International Conference*. SAS Institute Inc.; Cary, NC: 2005. Paper 201–230
 35. Wogan GN, Paglialunga S, Newberne PM. Carcinogenic effects of low dietary levels of Aflatoxin B1 in rats. *Food and Cosmetics Toxicology* 1974;12:681–685. [PubMed: 4375655]

Figure 1. Estimated extra risk (solid curve) and large-sample Scheffé-style 95% simultaneous confidence band (dashed curve) for carbendazim data.

Table 1
A summary of the Abbott-adjusted models used in the simulation study. For all models, $d_{\max}$ is set to 1.0.

Parameterizations	$R(0)$	$R(d_{\max})$	Weibull Model			Log-logistic Model		
			β_0	β_1	β_2	β_0	β_1	β_2
A	0.01	0.10	0.01	-2.35	1.63	0.01	-2.30	1.68
B	0.01	0.20	0.01	-1.55	1.76	0.01	-1.44	1.88
C	0.10	0.30	0.10	-1.38	1.59	0.10	-1.25	1.76
D	0.05	0.50	0.05	-0.44	1.08	0.05	-0.11	1.33
E	0.30	0.75	0.30	0.03	1.45	0.30	0.59	1.98
F	0.10	0.90	0.10	0.79	1.91	0.10	2.08	3.32

Table 2
Empirical coverage rates for the Scheffé-style upper band applied to Abbott-adjusted Weibull models (rates based on 2000 simulated data sets, nominal $\alpha = 0.05$). Model codes are from Table 1.

Model	N=25	N=50	N=100	N=300	N=500	N=1000	N=5000	N=10,000
A	1.0000	1.0000	1.0000	0.9960	0.9870	0.9855	0.9720	0.9690
B	1.0000	0.9995	0.9970	0.9835	0.9835	0.9690	0.9590	0.9540
C	1.0000	1.0000	1.0000	0.9965	0.9915	0.9855	0.9700	0.9670
D	0.9995	0.9960	0.9880	0.9740	0.9725	0.9700	0.9550	0.9580
E	1.0000	0.9995	0.9945	0.9870	0.9780	0.9680	0.9655	0.9610
F	0.9850	0.9795	0.9715	0.9630	0.9655	0.9600	0.9580	0.9575

Table 3
Empirical coverage rates for the Scheffé-style upper band applied to Abbott-adjusted log-logistic models (rates based on 2000 simulated data sets, nominal $\alpha = 0.05$). Model codes are from Table 1.

Model	N=25	N=50	N=100	N=300	N=500	N=1000	N=5000	N=10,000
A	1.0000	1.0000	0.9995	0.9985	0.9920	0.9850	0.9675	0.9575
B	1.0000	0.9965	0.9935	0.9795	0.9710	0.9645	0.9615	0.9525
C	1.0000	1.0000	1.0000	0.9905	0.9920	0.9835	0.9620	0.9645
D	0.9975	0.9920	0.9820	0.9695	0.9865	0.9570	0.9465	0.9530
E	1.0000	0.9975	0.9865	0.9755	0.9715	0.9570	0.9640	0.9500
F	0.9755	0.9600	0.9675	0.9605	0.9555	0.9565	0.9470	0.9480