Risk Analysis, Vol. 14, No. 6, 1994

Estimating Upper Confidence Limits for Extra Risk in Quantal Multistage Models

A. John Bailer,[1,2] and Randall J. Smith[2]

Received October 22, 1993; revised June 1, 1994

Multistage models are frequently applied in carcinogenic risk assessment. In their simplest form, these models relate the probability of tumor presence to some measure of dose. These models are then used to project the excess risk of tumor occurrence at doses frequently well below the lowest experimental dose. Upper confidence limits on the excess risk associated with exposures at these doses are then determined. A likelihood-based method is commonly used to determine these limits. We compare this method to two computationally intensive "bootstrap" methods for determining the 95% upper confidence limit on extra risk. The coverage probabilities and bias of likelihood-based and bootstrap estimates are examined in a simulation study of carcinogenicity experiments. The coverage probabilities of the nonparametric bootstrap method fell below 95% more frequently and by wider margins than the better-performing parametric bootstrap and likelihood-based methods. The relative bias of all estimators are seen to be affected by the amount of curvature in the true underlying dose–response function. In general, the likelihood-based method has the best coverage probability properties while the parametric bootstrap is less biased and less variable than the likelihood-based method. Ultimately, neither method is entirely satisfactory for highly curved dose–response patterns.

KEY WORDS: Dose–response models; bootstrapping; likelihood-based confidence intervals.

1. INTRODUCTION

Many challenges are encountered in the evaluation of the risks associated with exposure to potentially hazardous chemicals. One such challenge is the modeling of the risk of a deleterious response as a function of toxin dose. Data used in modeling risk often are from animal studies, with tumor onset in various tissues frequently used as the toxic end point. These data are used to fit a functional form relating the probability of toxic effect to dose which is used to estimate the extra risk associated with exposure to low doses of the chemical. Often an upper confidence limit is computed for this risk. Confidence levels of 95% are commonly used.

In this paper we discuss various methods for constructing upper-confidence level estimates of extra risk at a specified dose. We focus on the commonly used likelihood-based method and on computationally intensive bootstrap methods. These methods are discussed in Section 2. In Section 3, we compare these two methods using data from a recent long-term animal carcinogenicity experiment. We describe a simulation study of these upper-confidence limit estimators in Section 4.

2. METHODS

2.1 Data

Modeling tumor incidence minimally requires at least two pieces of information for each animal: expo-

[1] Department of Mathematics & Statistics, Miami University, Oxford, Ohio 45056.
[2] Risk Assessment Program, Division of Standards Development and Technology Transfer, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226.

0272-4332/94/1200-1001$07.00/1

sure and tumor status for each animal under study. With this information, simple procedures can be implemented to estimate model parameters. When additional information such as survival time and tumor context (incidental vs fatal tumors) is available, more complicated time-to-tumor analyses become viable options (cf. Ref. 2).

2.2 Quantal Multistage Models

Quantal multistage models relate the proportion of tumor-bearing animals in each dose group to an exponential model containing a polynomial in dose. These models were derived from early mechanistic models of carcinogenicity.[1] A common representation of this model is

$$P(d) = 1 - \exp(-q_0 - q_1 d - q_2 d^2 - \dots - q_k d^k)$$

where $P(d)$ is the probability of tumor onset in an animal exposed to dose d of a toxin and $(q_0, q_1, q_2, \dots, q_k)$ are nonnegative parameters. The degree of the polynomial is frequently set to be one less than the number of dose groups under study. The degree of the polynomial is also referred to as the number of stages affected by dose in the multistage model that motivated this functional form. The estimates for the model parameters are obtained using maximum-likelihood techniques.

Given the parameter estimates obtained from the fit of such a model, extrapolation to doses at levels of regulatory concern is often of interest. Two excess risk end points are commonly considered. The added risk (AR) associated with some d is the additional proportion of tumor incidence over background incidence in animals exposed to a dose of interest [(i.e., $\mathrm{AR}(d)=P(d)-P(0)$)]. Extra risk (ER) associated with a specified dose d is $\mathrm{ER}(d) = [P(d)-P(0)]/[1-P(0)]$, which represents the excess proportion of tumors among those animals that would have been tumor-free in the absence of exposure to the toxic dose. Point estimates of either risk end point may be based on using the estimated quantal multistage model, i.e., estimates ($\hat{q}$'s) are substituted for the parameters (q's) in the $P(d)$ function. These point estimates do not reflect the statistical variability in the data, and thus, confidence limits, typically upper confidence limits on risk, are also determined. The remainder of this paper focuses on upper confidence limit estimation for extra risk at a specified dose. (Note that the lower bound estimate on the dose associated with specified added or extra risk, the so-called "virtually safe dose," or VSD estimate, is also sometimes used in risk assessment.)

2.3. Upper-Confidence Limit Calculation

2.3.1. *Likelihood-Based Procedures*

Confidence interval calculation for risk in the context of quantal multistage models has been well studied,[4] with a good review of methods given by Crump and Howe.[5] Crump and Howe[5] reviewed the construction of confidence intervals based on distributional properties of maximum-likelihood estimates, distributional properties of likelihood ratios, and the bootstrap. They focused on estimating a lower limit for the dose associated with a specified level of extra risk (i.e., the VSD). The first two procedures were based on statistical properties associated with maximum-likelihood estimation (cf. Ref. 3). Crump and Howe[5] concluded that the likelihood ratio-based method was preferable to the method based on maximum-likelihood estimates for theoretical reasons (e.g., invariance under transformations) as well as their practical experience in low-dose extrapolation. Though they described bootstrapping as a method for constructing confidence intervals, bootstrapping was not considered in their simulation study. Crump and Howe[5] concluded their review by noting that "the bootstrap approach may be useful in low dose extrapolation, and further investigation into this method could be worthwhile" (p. 202).

Crump and Howe[5] noted potential difficulties that may arise from using likelihood-based methods. For example, if one (or more) of the parameters in the multistage model is zero (i.e., falls on the boundary of the parameter space), then confidence intervals based on features associated with the likelihood may not be valid. Crump *et al.*[4] compared the behavior of the likelihood methods with simulation-based "envelope curves"—essentially parametric bootstrap estimates of the added risk (see below). They observed that for the cases they studies, the "asymptotic [likelihood] confidence curve is either very close to or beyond the simulated envelope curve [which] indicates that the asymptotic confidence intervals may be somewhat conservative" (Ref. 4, p. 444). ("Conservative" in this context means that the confidence interval coverage probability exceeds the nominally specified confidence level.) Crump *et al.*[4] used these "envelope curves" as a standard for studying the behavior of the likelihood methods and proposed these curves as an alternative to the likelihood methods.

2.3.1. Bootstrap Procedures

Bootstrap procedures are computationally intensive methods for generating an estimate of the sampling distribution of a statistic that can be used in confidence interval construction.[7] Bootstrap methods involve using the observed data to simulate the experiment a large number of times. The statistic of interest (e.g., extra risk at some specified dose) is calculated for each of the simulated experiments, and then the 95th percentile of these statistics is used to obtain the 95% upper confidence limit. This bootstrap procedure for constructing confidence intervals is sometimes referred to as the "percentile" method.[7]

Quantal multistage models are fit to data from studies that can be conceptualized as a series of separate experiments. Each experiment is characterized by a dose level (d_i), the number of animals at risk of tumor onset (n_i), and the probability of tumor onset (p_i). The outcome of each experiment is the observed number of tumor-bearing animals (x_i), which can be viewed as a binomally distributed random variable with parameters n_i and p_i. The simulated experiments used in the bootstrap confidence procedure simply mimic these binomial experiments—i.e., for dose group i, a binomial random variable (x_i*) with parameters n_i and p_i is generated. Two common choices for the probability parameter (p_i) are x_i/n_i (the "nonparametric bootstrap") and $\hat{P}(d_i)$ (the "parametric bootstrap"), where $\hat{P}(d_i)$ is based on the quantal multistage model, in which the maximum-likelihood estimates are substituted for the parameters (q's). The nonparametric bootstrap uses the observed proportion of tumor-bearing animals as an estimate of p_i. If no tumor-bearing animals are observed in a particular dose group (x_i=0), then the binomial experiment using x_i/n_i would always generate a zero response for that group. This could potentially cause the nonparamtric bootstrap method to be less variable than the parametric bootstrap.

Smith and Sielkin[12] continued the study of VSD estimation via bootstrap methods. They concluded that a "simple bootstrap procedure offers improvements over the current likelihood-ratio-based confidence limit procedure with virtually no undesirable side-effects" (p. 172), due to better coverage probability properties. Within the summary tables presented by Smith and Sielkin, there were indications of conservative behavior for cases with curvature and some anticonservative indications for other conditions. Finally, this study considered only a limited number of simulation conditions. Namely, only conditions with a zero background tumor rate and a fixed tumor count in the high-dose group (at 30 tumors of 50 at risk) were simulated. These limitations removed two potential sources of variability from the bootstrap samples.

Even with the warnings associated with the use of likelihood methods, these techniques are frequently the default selection for representing statistical variability in risk estimates. To illustrate this potential problem, we compare the behavior of likelihood ratio-based and bootstrap-based methods for calculating upper confidence limits on risk at specified doses. Two simple data sets demonstrate the potential discrepancies in risk estimates based on these methods.

3. ILLUSTRATION AND MOTIVATION

We compare the estimated upper confidence limits on risk associated with low-dose exposure to 1,3-butadiene. This chemical has been studied in a long-term animal carcinogenicity experiment in which male and female B6C3F$_1$ mice were exposed to 1,3-butadiene concentrations of 0, 6.25, 20, 62.5, 200, or 625 ppm.[9] Statistically significant increases in tumor onset were observed in six different sites in male mice and eight different sites in female mice. These data were recently used on the basis for a risk assessment.[6] We use the observed tumor onset in lung adenoma/carcinomas and in heart hemangiosarcomas in female mice to illustrate differences in the behavior of bootstrap-based versus likelihood-based upper confidence limit calculations of extra risk. Data for these two sites along with parameter estimates of multistage model fits are displayed in Table I. (Though this experiment was conducted at exposure levels up to 625 ppm, only the lowest dose groups for which a quantal multistage model adequately fit the data are included in Table I and the accompanying analysis.) From Table I, we see an observed tumor pattern for heart hemangiosarcomas which is sublinear, having no tumors in the control and lowest two concentration groups. The pattern of lung tumor onset appears to be linear, with increasing tumor incidence associated with increasing 1,3-butadiene concentration. The fits of the quantal multistate model to these data are consistent with the observations given above. The estimate of the linear parameter in the multistage model for the heart hemangiosarcoma data was zero, while the model fit to the lung data had zero estimates associated with the quadratic and cubic terms.

For purposes of comparison, 95% upper confidence limits on extra risk were calculated for concentrations of 2, 0.2, and 0.02 ppm. The likelihood estimates for extra risk were calculated using GLOBAL83.[8] Both nonparametric and parametric bootstrap estimates were calculated based on 1000 bootstrap samples. The bootstrap estimates were calculated using a FORTRAN program running on an HP 9000/720 computer. The results of these confidence limit calculation procedures are presented in Table II. For the tumor incidence data in the lung, all procedures yielded approximately the same upper confidence limit at each of the target concentrations. However, for the hemangiosarcoma data, the likelihood-based confidence limit is much larger than the bootstrap confidence limit, with the discrepancy increasing as the target concentration decreases.

Given that we do not know the underlying dose–response pattern and hence the true value of extra risk, it is not clear which method is more correct. At this point, the extra risk estimates appear to be strongly influenced both by the degree of curvature in the dose–response pattern and by the extent of the low-dose extrapolation. A simulation study was conducted to explore these questions.

4. SIMULATION STUDY

4.1. Description

A simulation study was constructed to explore the behavior of various methods for estimating 95% upper confidence limits for extra risk. The study consisted of repeatedly generating 1000 carcinogenicity experiments for each of 16 conditions, followed by computation of the upper confidence limits using nonparametric and parametric bootstrap procedures and the likelihood-based procedure. One thousand bootstrap samples were generated for each simulated carcinogenicity experiment for each of the nonparametric and parametric estimation procedures. The simulated carcinogenicity study followed the protocol used in most recent National Toxicology Program (NTP) studies. In these studies, four dose groups, with 50 animals in each group, are spaced at doses (d) of 0, $\frac{1}{4}$, $\frac{1}{2}$, and 1 MTD (the so-called "maximum tolerated dose"). In a four-group study, quantal models up to three stages were considered, i.e.,

Table I. Tumor Data and Multistage Model Parameter Estimates[a] for Female B6C3F$_1$ Mice Exposed to 1,3-Butadiene

Site (parameter estimates)	Dose (ppm)	Number of tumor-bearing animals (number at risk)
Heart hemangiosarcomas	0	0 (50)
$\hat{q}_0 = \hat{q}_1 = 0$	6.25	0 (50)
$\hat{q}_2 = 4.416e{-}7$	20	0 (50)
$\hat{q}_3 = 6.620e{-}8$[b]	62.5	1 (49)
	200	21 (50)
Lung adenomas & carcinomas	0	4 (50)
$\hat{q}_0 = 0.181$	6.25	15 (50)
$\hat{q}_1 = 9.554e{-}3$;	20	19 (50)
$\hat{q}_2 = \hat{q}_3 = 0$	62.5	24 (50)

[a] Parameter estimates arise from fitting the model $P(d) = 1 - \exp(-q_0 - q_1d - q_2d^2 - q_3d^3)$ to these data sets (see text for greater detail).
[b] The notation "$ae-b$" is used to represent the quantity "$a \times 10^{-b}$."

$$P(d) = 1 - \exp(-q_0 - q_1d - q_2d^2 - q_3d^3)$$

Specification of the parameters of this model determined the underlying dose response patterns.

Four background tumor rates (1, 5, 10, and 30%) were considered. These rates roughly corresponded to the range of background tumor rates observed in NTP studies (see, e.g., Ref. 10). Additionally, the tumor response at the highest dose group was set equal to 8, 15, 30, and 90% for background tumor rates of 1, 5, 10, and 30%, respectively. Finally, four levels of curvature were considered:

L: Linear—$q_2=q_3=0$
I1: Low curvature 1—All q's > 0
I2: Moderate curvature 2—All q's > 0
C: High curvature—$q_1=q_2=0$

The 4 (background) × 4 (curvature) = 16 conditions are displayed in Fig. 1, with coefficients given in Table III. As noted above, 1000 simulated experiments were generated for each of these conditions. This number of simulated experiments provided a margin of error of 1.35% for estimating coverage probabilities associated with a 95% nominal coverage level. (Aside: Each of these simulation conditions took approximately 22.5 CPU h using a HP/Apollo 9000/720.)

Three multistage models were fit to each simulated data set, with confidence limits for extra risk calculated for each model using each procedure. The multistage models that were fit included the following:

1s: One stage—$P(d) = 1 - \exp(-q_0 - q_1d)$

Table II. Upper 95% Confidence Limits (UCL) on Extra Risk Based on Likelihood-Ratio and Bootstrap Methods

			Upper 95% confidence limit		Ratio of UCL estimates (likelihood/Bootstrap[a])	
			Bootstrap[a]		Bootstrap[a]	
Site	Dose (ppm)	Likelihood	NP	P	NP	P
Heart hemangiosarcomas	2.00	1.404e−3[b]	5.167e−5	5.168e−5	27.2	27.2
	.20	1.405e−4	5.154e−7	5.156e−7	272.5	272.5
	.02	1.405e−5	5.153e−9	5.155e−9	2726.3	2725.5
Lung	2.00	2.835e−2	2.883e−2	2.628e−2	1.0	1.1
	.20	2.872e−3	2.921e−3	2.659e−3	1.0	1.1
	.02	2.876e−4	2.925e−4	2.662e−4	1.0	1.1

[a] "NP" ("P") corresponds to the nonparametric (parametric) bootstrap estimate.
[b] The notation "ae−b" is used to represent the quantity "$a \times 10^{-b}$."

2s: Two stage—
$P(d) = 1 - \exp(-q_0 - q_1 d - q_2 d^2)$
3s: Three stage—
$P(d) = 1 - \exp(-q_0 - q_1 d - q_2 d^2 - q_3 d^3)$

The fitting of quantal models with number of stages equal to one less than the number of dose groups is a common practice.

Extra risk estimation was conducted as each of three dose levels (0.1, 0.01, 0.001).

4.2. Results

4.2.1. *Simulation Validity Checks*

Various checks of the simulation experiment were conducted. Histograms of tumor counts in the four dose groups were compared to expected tumor counts based upon the underlying dose–response patterns. Histograms of parameter estimates were compared to the input values (q_0, q_1, q_2, q_3). Both of these comparisons provided support of the validity of the simulation experiment. Finally, the FORTRAN simulation coding was checked by comparison of fits with a standard package (GLOBAL83). The FORTRAN program generated comparable estimates of the coefficients (q's) and likelihood-based UCLs as did GLOBAL83 for the test cases.

4.2.2. *Coverage Probabilities*

A coverage probability represents the proportion of times over repeated samples that an upper-confidence limit estimate of extra risk exceeds the true extra risk. The nominal value of this quantity is specified when constructing such limits. One measure of the quality of a statistical procedure is that its actual coverage probability is approximately equal to the nominally stated level.

Estimated coverage probabilities for 95% upper confidence level extra risk estimates (UCL) are presented in Tables IV and V for the three methods (N, nonparametric bootstrap; P, parametric bootstrap; LR, likelihood-ratio based). Each table displays the results for four background tumor rates (1, 5, 10, 30) and four levels of curvature in dose–response (L, I1, I2, C). Table IV shows the results for the three doses (0.1, 0.01, 0.001) when fitting the three-stage model. Table V shows the results from fitting the three quantal models involving different numbers of stages (1s, 2s, 3s) at dose=0.01.

As shown in Table IV, the nonparametric bootstrap is frequently anticonservative (does not attain the nominally stated coverage probability) for the 3s model fits. In contrast, the parametric and likelihood-based procedures are frequently conservative (exceeding the nominally stated coverage probability). However, a notable exception is the "L" (linear) dose–response pattern, where both parametric and likelihood-based procedures were anticonservative as well. Anticonservative results from the likelihood-based method were unexpected since it is reputed to be conservative. For all procedures, the coverage probabilities increased as the degree of curvature increased.

As shown in Table V, coverage probabilities tended to decrease as the number of stages in the fitted model increased. This is not surprising since 1s models contain only a linear dose term which would bound above any sublinear pattern, leading to relatively larger coverage

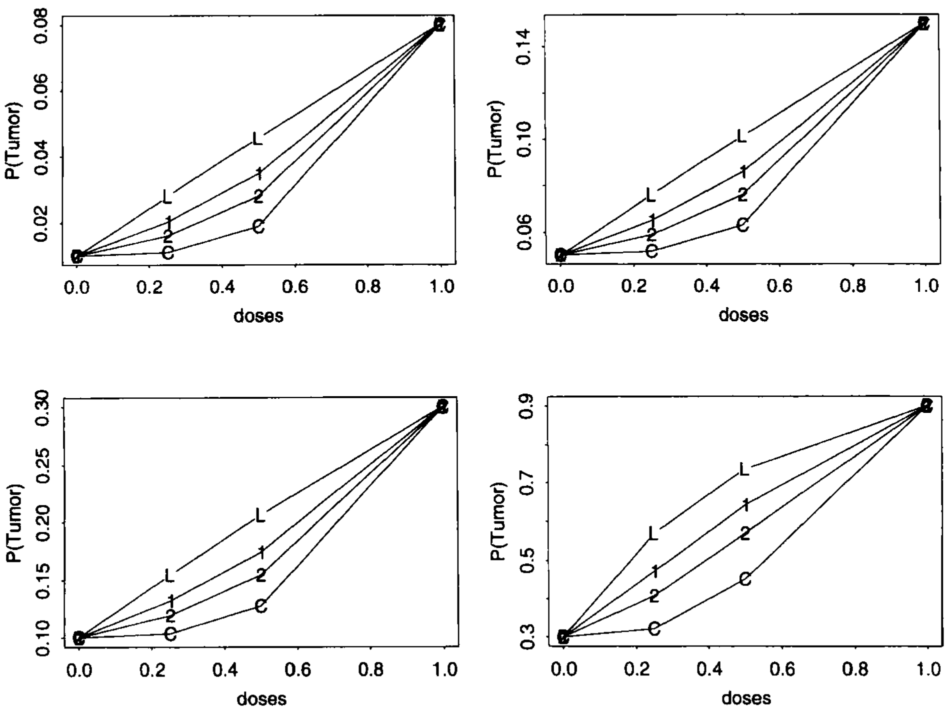

Fig. 1. Plots representing 16 simulation conditions displaying four background tumor rates (1%—row 1, column 1; 5%—row 1, column 2; 10%—row 2, column 1; 30%—row 2, column 2) and four levels of curvature in dose–response (L, linear; 1, intermediate curvature 1; 2, intermediate curvature 2; C, high curvature).

probabilities than models that would allow for greater degrees of curvature.

4.2.3. UCL Relationship to True Extra Risk

In addition to coverage probabilities, we compared the methods in terms of how far the estimated UCL for extra risk ($\hat{ER}_{UCL}$) was from the true extra risk (ER). Table VI displays the results for average relative bias. These results were calculated by averaging the 1000 values of ($\hat{ER}_{UCL}$−ER)/ER over simulated experiments for each method. The relative bias should exceed 0 since $\hat{ER}_{UCL}$ is an UCL on extra risk. However, among two upper-confidence limit estimators that both maintained nominal coverage probabilities, the estimator that is relatively closer to the true extra risk might be preferred. The standard deviation of the ($\hat{ER}_{UCL}$−ER)/ER values also was presented in Table VI.

The results in Table VI suggest that the likelihood-based procedure tends to be relatively farther above the true extra risk and more variable than the bootstrap procedures. Second, the distance from true extra risk decreases as the background tumor rates increase. Finally, all three procedures tend to generate extra risk estimates that are relatively farther away from the true extra risk as the degree of curvature associated with the underlying dose–response function increases. For the underlying dose–response pattern that included only dose as a cubic term (the highest degree of curvature considered), the relative bias increased dramatically relative to the other simulation conditions. This increase became more pronounced as the extra risk was estimated at ever lower dose values.

5. DISCUSSION

In the motivating example, we saw that in a situation where a linear term is positive in a quantal multistage model, both likelihood- and bootstrap-based upper confidence level risk estimates were similar. But the site where the linear term in the quantal multistage model fit is zero, the upper-confidence level risk calculations differed dramatically between likelihood-based and boot-

Table III. Multistage Model Coefficients Specifying 16 Simulation Conditions Representing 4 Background Tumor Rates (1, 5, 10, 30%) and 4 Levels of Curvature in Dose–Response (L, linear; I1, intermediate curvature 1; I2, Intermediate Curvature 2; C, High Curvature)

		Degree of curvature			
Background	Coefficient	L	I1	I2	C
1%	q_0	0.01005	0.01005	0.01005	0.01005
	q_1	0.07333	0.03667	0.01833	0
	q_2	0	0.01833	0.01833	0
	q_3	0	0.01833	0.03667	0.07333
5%	q_0	0.05129	0.05129	0.05129	0.05129
	q_1	0.11123	0.05561	0.02781	0
	q_2	0	0.02781	0.02781	0
	q_3	0	0.02781	0.05561	0.11123
10%	q_0	0.10536	0.10536	0.10536	0.10536
	q_1	0.25131	0.12566	0.06283	0
	q_2	0	0.06283	0.06283	0
	q_3	0	0.06283	0.12566	0.25131
30%	q_0	0.35667	0.35667	0.35667	0.35667
	q_1	1.94592	0.97296	0.48648	0
	q_2	0	0.48648	0.48648	0
	q_3	0	0.48648	0.97296	1.94592

strap-based estimates. The discrepancy between the bootstrap- and the likelihood-ratio methods is a cause for concern and suggests that these methods may have had considerably different actual confidence levels. Hence, one and possibly all methods may not maintain the nominally stated confidence level. One possible explanation is that the likelihood procedure leads to an overly conservative risk estimate for sublinear dose–response patterns due to its inherent linear behavior in low-dose risk estimation. These potential explanations were explored with a simulation study.

Observations based on this study include the following.

- Even a so-called conservative estimation procedure, the likelihood-based method, can be anticonservative in certain nonpathological situations, namely, a linear model. All three procedures are anticonservative in this situation, with the likelihood-based method closest to the nominal coverage probability.
- The nonparametric bootstrap percentile method is not a viable option for situations in which tumors also occur in control conditions due to poor coverage probability properties (As an aside, a simple modification of the nonparametric bootstrap in which some pooling of data from adjacent dose groups where monotonic increases in tumor burden were violated prior to bootstrapping might improve the coverage probability properties of the nonparametric estimator.)
- The 95% UCLs for extra risk are conservative for dose–response relationships truly possessing a high curvature (in some cases, 1,000,000% above true extra risk).

In the simulation study presented herein, the situation illustrated in the Section 3 example was not reproduced. This highlights a potential shortcoming associated with all simulation studies. Conclusions and generalizations are naturally constrained by the conditions simulated. We attempted to select conditions that spanned a broad range of background tumor rates, dose–response patterns, low-dose extrapolation conditions, and stages of models fit.

In order to explore the difference between UCL estimation methods highlighted in the 1,3-butadiene example, we simulated an additional experimental situation with a zero background tumor rate (q_0=0) and high curvature (as in the "C" condition). As in the C condition, all UCL methods exceeded the nominal coverage probability. However, the nonparametric bootstrap average UCL was appreciably closer to the true ER relative to both the parametric bootstrap- and the likelihood-based average UCLs. As reported by others,[11,12] we believe that this reflects the importance of the parameter associated with the linear dose term (q_1) on UCL estimation. If a nonzero q_1 estimate occurs at a frequency greater than 5% in such highly curved dose–response data sets, then 95% UCLs will be conservative. Based upon these simulations, we conjecture that the bootstrap methods will be preferable to the likelihood-based method for UCL estimation for situations in which a large number of dose-group data are available (say more than four groups), with many of the lower dose groups exhibiting zero tumor counts.

In conclusion, the parametric bootstrap-based procedure deserves more attention as a means of generating upper confidence limit estimates for extra risk in low-dose regions. This method has coverage probability properties similar to those of the likelihood-based method under most of the simulation conditions while being slightly closer to true extra risk and less variable. Ultimately, none of these procedures adequately dealt with the condition of high curvature. This may reflect an inherent difficulty with using estimation techniques that revolve around the use of the quantal multistage model.

Table IV. Estimated Coverage Probabilities for 95% Upper-Confidence Level Extra Risk Estimates[a]

	0.1				0.01				0.001			
	L	I1	I2	C	L	I1	I2	C	L	I1	I2	C
1%												
N	0.76	0.84	0.86	0.95	0.76	0.84	0.86	0.95	0.76	0.84	0.85	0.95
P	0.82	0.97	0.99	1	0.82	0.97	0.99	1	0.82	0.97	0.99	1
LR	0.89	0.99	1	1	0.89	0.99	1	1	0.89	0.99	1	1
5%												
N	0.85	0.90	0.92	0.99	0.84	0.91	0.91	0.99	0.84	0.91	0.91	0.99
P	0.90	1	1	1	0.90	1	1	1	0.90	1	1	1
LR	0.91	0.99	1	1	0.91	0.99	1	1	0.91	0.99	1	1
10%												
N	0.86	0.92	0.93	1	0.85	0.92	0.93	1	0.85	0.92	0.93	1
P	0.89	0.99	1	1	0.88	0.99	1	1	0.88	0.99	1	1
LR	0.91	0.99	1	1	0.91	0.99	1	1	0.91	0.99	1	1
30%												
N	0.86	0.95	0.90	1	0.86	0.95	0.90	1	0.86	0.95	0.90	1
P	0.86	0.96	0.99	1	0.86	0.96	1	1	0.86	0.96	1	1
LR	0.90	0.97	0.96	1	0.90	0.97	0.97	1	0.90	0.98	0.97	1

[a] Simulation results for three methods of extra risk estimation (N, nonparametric bootstrap; P, parametric bootstrap; LR, likelihood-ratio based) are presented for four background tumor rates (1, 5, 10, 30), four levels of curvature in dose–response (L, I1, I2, C), and three doses (0.1, 0.01, 0.001). All estimates are based upon fitting a so-called three-stage model.

Table V. Estimated Coverage Probabilities for 95% Upper-Confidence Level Extra Risk Estimates[a]

	L			I1			I2			C		
	1s	2s	3s	1s	2s	3s	1s	2s	3s	1s	2s	3s
1%												
N	0.89	0.76	0.76	0.98	0.85	0.84	0.99	0.89	0.86	0.99	0.99	0.95
P	0.90	0.84	0.82	0.99	0.98	0.97	1	1	0.99	1	1	1
LR	0.93	0.90	0.89	0.99	0.99	0.99	1	1	1	1	1	1
5%												
N	0.93	0.85	0.84	0.99	0.90	0.91	1	0.93	0.91	1	1	0.99
P	0.94	0.91	0.90	1	1	1	1	1	1	1	1	1
LR	0.94	0.93	0.91	0.99	0.99	0.99	1	1	1	1	1	1
10%												
N	0.95	0.86	0.85	1	0.92	0.92	1	0.93	0.93	1	1	1
P	0.95	0.90	0.88	1	0.99	0.99	1	1	1	1	1	1
LR	0.95	0.92	0.91	1	1.00	0.99	1	1	1	1	1	1
30%												
N	0.96	0.86	0.86	1	0.92	0.95	1	0.84	0.90	1	1	1
P	0.96	0.86	0.86	1	1.00	0.96	1	1	1	1	1	1
LR	0.95	0.90	0.90	1	0.94	0.97	1	0.97	0.97	1	1	1

[a] Simulation results for three methods of extra risk estimation (N, nonparametric bootstrap; P, parametric bootstrap; LR, likelihood-ratio based) are presented for four background tumor rates (1, 5, 10, 30), three levels of curvature in dose–response (L, I1, I2, C), and three quantal models (1s, 2s, 3s). All estimates are based upon estimating extra risk at dose=0.01.

Table VI. Estimated Relative Bias (SE) Associated with 95% Upper-Confidence Level Extra Risk Estimates $\left[\text{Relative Bias} = \text{Mean of } \frac{(\hat{ER}_{UCL} - ER)}{ER}\right]^a$

	0.1				0.01				0.001			
	L	I1	I2	C	L	I1	I2	C	L	I1	I2	C
1%												
N	0.52	1.40	2.71	76.9	0.51	1.51	3.06	7,618	0.51	1.52	3.10	7.6e+5
	(0.72)	(1.27)	(2.29)	(60.3)	(0.74)	(1.36)	(2.62)	(6,199)	(0.74)	(1.37)	(2.65)	(6.2e+5)
P	0.57	1.62	3.39	104	0.57	1.75	3.86	10,494	0.57	1.77	3.90	1.0e+6
	(0.60)	(0.98)	(1.75)	(44)	(0.61)	(1.04)	(1.96)	(4,460)	(0.61)	(1.05)	(1.98)	(4.5e+5)
LR	0.81	2.03	4.05	120	0.81	2.20	4.62	12,146	0.82	2.21	4.68	1.2e+6
	(0.67)	(1.13)	(2.00)	(52)	(0.67)	(1.20)	(2.24)	(5,271)	(0.60)	(1.20)	(2.27)	(5.3e+5)
5%												
N	0.73	1.83	3.98	120	0.74	1.97	4.52	12,061	0.74	1.99	4.57	1.2e+6
	(0.70)	(1.33)	(2.47)	(67)	(0.70)	(1.43)	(2.82)	(6,872)	(0.71)	(1.44)	(2.85)	(6.9e+5)
P	0.75	2.07	4.55	146	0.75	2.23	5.18	14,745	0.75	2.25	5.24	1.5e+6
	(0.56)	(0.96)	(1.70)	(42)	(0.57)	(1.02)	(1.91)	(4,264)	(0.57)	(1.03)	(1.93)	(4.3e+5)
LR	0.89	2.23	4.80	147	0.90	2.42	5.48	14,894	0.90	2.44	5.55	1.5e+6
	(0.63)	(1.16)	(2.12)	(56)	(0.64)	(1.24)	(2.39)	(5,653)	(0.64)	(1.24)	(2.42)	(5.7e+5)
10%												
N	0.48	1.36	2.97	85.7	0.48	1.49	3.42	8,636	0.49	1.50	3.46	8.6e+5
	(0.47)	(0.90)	(1.76)	(50)	(0.48)	(0.97)	(2.03)	(5,219)	(0.48)	(0.98)	(2.06)	(5.2e+5)
P	0.46	1.48	3.41	109	0.46	1.62	3.93	11,099	0.46	1.64	3.98	1.1e+6
	(0.39)	(0.66)	(1.17)	(27)	(0.40)	(0.71)	(1.33)	(2,784)	(0.41)	(0.71)	(1.35)	(2.8e+5)
LR	0.57	1.62	3.55	105	0.58	1.78	4.10	10,742	0.58	1.88	4.16	1.1e+6
	(0.43)	(0.78)	(1.51)	(41)	(0.44)	(0.84)	(1.72)	(4,445)	(0.44)	(0.85)	(1.74)	(4.3e+5)
30%												
N	0.20	0.64	1.10	25.5	0.22	0.76	1.36	2,525	0.23	0.78	1.39	2.5e+5
	(0.19)	(0.38)	(0.79)	(17.8)	(0.21)	(0.45)	(0.98)	(1,976)	(0.22)	(0.46)	(1.00)	(2.0e+5)
P	0.18	0.63	1.18	34.6	0.20	0.76	1.46	3,550	0.21	0.77	1.49	3.6e+5
	(0.18)	(0.36)	(0.67)	(10.7)	(0.20)	(0.43)	(0.82)	(1,163)	(0.21)	(0.43)	(0.84)	(1.2e+5)
LR	0.23	0.72	1.28	32.3	0.26	0.87	1.62	3,380	0.26	0.89	1.65	3.4e+5
	(0.18)	(0.37)	(0.73)	(14.8)	(0.20)	(0.44)	(0.89)	(1,602)	(0.21)	(0.45)	(0.91)	(1.6e+5)

[a] Simulation results for three methods of extra risk estimation (N, nonparametric bootstrap; P, parametric bootstrap; LR, likelihood-ratio based) are presented for four background tumor rates (1, 5, 10, 30), four levels of curvature in dose–response (L, I1, I2, C), and three doses (0.1, 0.01, 0.001). All estimates are based upon fitting a so-called three-stage model.

ACKNOWLEDGMENTS

We would like to thank Dave Dankovic, Leslie Stayner, and Steve Gilbert for early discussions regarding upper-confidence limit risk estimation. We would like to acknowledge Christopher Portier for his suggestions regarding potential improvements to the performance of the nonparametric bootstrap estimator and for suggestions regarding the computer implementation of the likelihood-based method. Finally, we would like to thank Ralph Kodell, and Laurence Reed for their helpful comments based on a careful reading of an early version of the manuscript.

REFERENCES

1. P. Armitage and R. Doll, "Stochastic Models for Carcinogenesis," in J. Neyman (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4* (University of California Press, Berkeley and Los Angeles, 1961), pp. 19–38.
2. C. Brown and J. Koziol, "Statistical Aspects of the Estimation of Human Risk from Suspected Environmental Carcinogens," *SIAM Rev.* **25**, 151–181 (1983).
3. D. R. Cox and D. V. Hinkley, *Theoretical Statistics* (Chapman and Hall, London, 1974).
4. K. S. Crump, H. A. Guess, and K. L. Deal, "Confidence Intervals and Test of Hypotheses Concerning Dose Response Relations Inferred from Animal Carcinogenicity Data," *Biometrics* **33**, 437–451 (1977).
5. K. S. Crump and R. B. Howe, "A Review of Methods for Cal-

culating Statistical Confidence Limits in Low Dose Extrapolation," in D. B. Clayson, D. Krewski, and I. Munro (eds.), *Toxicological Risk Assessment, Vol. 1* (CRC Press, Boca Raton, FL, 1985).
6. D. Dankovic, R. Smith, L. Stayner, and A. J. Bailer, "Time-to-Tumor Risk Assessment for 1,3-Butadiene Based on Exposure of Mice to Low Doses by Inhalation," in M. Sorsa, K. Peltonen, H. Vainio and K. Hemminki, (eds.), *Butadiene and Styrene: Assessment of Health Hazards* (IARC Scientific Publications No. 127, Lyon, 1993), pp. 335–344.
7. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (SIAM, Philadelphia, 1982).
8. R. B. Howe and K. S. Crump, *GLOBAL 83* (Science Research Systems, Ruston, LA, 1983).
9. National Toxicology Program (NTP), "Toxicology and Carcinogenesis Studies of 1,3-Butadiene (CAS No. 106-99-0) in B6C3F1 Mice (Inhalation Studies), NTP TR 434," NIH Publication No. 93-3165 (National Toxicology Program, Research Triangle Park, NC, 1993).
10. C. Portier, J. Hedges, and D. Hoel, "Age-Specific Models of Mortality and Tumor Onset for Historical Control Animals in the National Toxicology Program's Carcinogenicity Experiments," *Cancer Res.* **46**, 4372–4378 (1986).
11. C. Portier and D. Hoel, "Lose-Dose-Rate Extrapolation Using the Multistage Model," *Biometrics* **39**, 897–906 (1983).
12. L. A. Smith and R. L. Sielkin, "Bootstrap Bounds for "Safe" Doses in the Multistage Cancer Dose-Response Model," *Commun. Stat. Simul.* **17**, 153–175 (1988).