Model Uncertainty and Model Averaging in the Estimation of Infectious Doses for Microbial Pathogens

Hojin Moon[1], Steven B. Kim[2], James J. Chen[3], Ralph L. Kodell[4]
[1]Department of Mathematics and Statistics, California State University, Long Beach, CA 90840-1001
[2]Department of Statistics, University of California, Irvine, CA 92697
[3]Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079
[4]Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR 72205

Abstract

Food-borne infection is caused by intake of foods or beverages contaminated with microbial pathogens. Dose-response modeling is used to estimate exposure levels of pathogens associated with specific risks of infection or illness. When a single dose-response model is used and confidence limits on infectious doses are calculated, only data uncertainty is captured. We propose a method to estimate the lower confidence limit on an infectious dose by including model uncertainty and separating it from data uncertainty. The infectious dose is estimated by a weighted average of effective dose estimates from a set of dose-response models via a Kullback information criterion. The confidence interval for the infectious dose is constructed by the delta method, where data uncertainty is addressed by a bootstrap method. To evaluate the actual coverage probabilities of the lower confidence limit, a Monte Carlo simulation study is conducted under sublinear, linear and superlinear dose-response shapes that can be commonly found in real data sets. Our model-averaging method achieves coverage close to nominal in almost all cases, thus providing a useful and efficient tool for accurate calculation of lower confidence limits on infectious doses.

Key Words: bias-skewness correction; confidence limit; data uncertainty; food safety; Kullback information criterion

1. Introduction

Food safety is a critical issue in public health. The Centers for Disease Control and Prevention (CDC, 2011) estimated that there are 48 million illnesses, 128 thousand hospitalizations, and 3000 deaths due to foodborne pathogens every year in the United States. Among well-known pathogens, *Norovirus* and *Salmonella* are included in the top five pathogens causing domestically acquired foodborne illness and resulting in hospitalization and death. Another well-known pathogen, *E. coli* O157*,* was included in the top five pathogens causing domestically acquired foodborne illnesses resulting in hospitalization.

In order to control diseases caused by microbial contaminants in food, it is essential to assess their dose-response relationships as accurately as possible. However, definitive dose-response data on humans at low levels of contamination likely to occur in practice are scarce to nonexistent. Hence, when sufficient animal or human data at high doses are available to allow dose-response modeling, allowable contamination levels of specific microorganisms in food can be derived using infectious dose (*ID*) levels derived from these models as "points of departure" for low-dose extrapolation. This approach to setting exposure levels is equivalent to the benchmark dose (BMD) approach used in chemical risk assessment (EPA, 2000). Reliable methods for deriving such ID_p levels ($0.01 \leq p \leq 0.10$) are essential, where ID_p is defined as a dose that causes a response (infection or illness) at a predetermined risk level, p.

Dose-response models with one, two, and three parameters have been proposed for dose-response modeling in microbial risk assessment (MRA) (Kodell et al, 2002; Marks et al, 1998; Moon et al., 2004). The simplest model is the one-parameter exponential model, $P(d;\ \alpha) = 1 - exp(-\alpha d), \alpha > 0,$ which can be derived from basic biological assumptions considering low numbers of pathogens as discrete particles (Haas, 1983). Even though the exponential model has low-dose linearity, it is often not flexible enough to provide an adequate fit to dose-response data on pathogenic microorganisms. The Beta-Poisson (BP) model (Haas, 1983; Furumoto and Mickey, 1967; Haas et al., 1999), which includes slightly more complex biology, has been used in MRA(WHO, 2001a, 2001b, 2002). However, the adequacy of the BP model as a potential "default" model for MRA has been questioned (Marks et al., 1998). It can be shown that the exponential and the BP models are dose-response pattern-specific so that they may not be suitable models under a certain dose-response pattern (e.g. a hypothetical sublinear pattern in Figure 1). Other two-parameter models include the Log-Normal (LN), the Log-Logistic (LL), and the Extreme-Value (EV) models (Pinsky, 2000). The best-known three-parameter model is the Weibull-Gamma (WG) model (Farber et al., 1996). However, three-parameter models require data at four or more dose levels, which may not be readily available for many microbial agents. It was noted that the BP and LL models are special cases of the WG model (Kodell et al, 2002).

Several dose-response models often provide reasonably good fits to the data in the experimental dose range but can yield very different infectious dose (*ID*) estimates in the low-dose range, even with infection rates as high as $0.01 \leq p \leq 0.10$. Hence, it is undesirable to choose only one model and estimate an *ID* based on the chosen model. In order to account for model uncertainty, model averaging (MA) methods have been proposed (FDA/FSIS, 2003; Moon et al., 2004, 2005; Bailer et al., 2005; Faes et al., 2007; Wheeler and Bailer, 2007; Namata et al., 2008).

Kang et al. (2000) used four two-parameter models to demonstrate how model uncertainty can be addressed in MRA using the Akaike information criterion (AIC) (Akaike, 1974) to average the individual-model *ID*s. Kodell et al. (2002) presented a general framework for generating dose-response models in the interest of deriving potential competitors for the three-parameter Weibull-Gamma (WG) model (Farber et al., 1996). Moon et al. (2004) suggested that two-parameter dose-response models for MRA reflected at least as much model uncertainty as three-parameter models. Moon et al. (2005) used maximum likelihood estimates (MLEs) from a binomial log-likelihood function to estimate parameters of dose-response models and integrated model uncertainty into estimating *ID*s using weights obtained from the Kullback information

criterion (Cavanaugh, 1999; Kim and Cavanaugh, 2005) as a measure of model variability.

Faes et al. (2007) showed the necessity of model averaging by comparing the *ID* based on a set of fractional polynomials to the *ID* from the selected best model. Namata et al. (2008) also investigated model averaging in MRA using fractional polynomials, suggesting that the common two-parameter models were not sufficiently diverse to give appropriate confidence coverage in model averaging.

Bayesian model averaging provides a coherent approach for accounting for model uncertainty (Hoeting et al., 1999). Bailer et al. (2005) illustrated Bayesian model averaging with a simple Bayesian information criterion (BIC) approximation (Schwarz, 1978; Kass and Wasserman, 1995) under the assumption of the unit information prior on the parameter space for addressing uncertainty in the selection of models when generating risk estimates. Their emphasis was on the Bayesian analysis of model uncertainty to obtain a model-averaged summary.

Recently, Wheeler and Bailer (2007) investigated an alternative MA approach to estimate *ID*s (benchmark doses, or BMDs, in their terminology) based on a weighted "average model" and illustrated their method with dose-response lung cancer data on rats. Their *ID* estimates are obtained by solving the weighted average model, and the weights are determined by AIC. The lower confidence bound on the *ID* was obtained by the bootstrap percentile method. They investigated the coverage of their *ID* lower confidence estimates with linear and sublinear dose-response patterns.

We propose a new method to estimate the lower confidence limit (LCL) for an *ID* under the consideration of both model uncertainty and data uncertainty. Following Moon et al. (2004), four two-parameter dose-response models are considered a sufficiently diverse set of models. Parameters of the models are estimated by the method of maximum likelihood. The *ID* is estimated by a KIC-weighted average of effective dose (ED) estimates from the dose-response models. We introduce a new variance formula and calculate the variance of the *ID* estimate with separate components for model uncertainty and data uncertainty via a bootstrap method. The LCL for the *ID* is constructed assuming the *ID* is log-normally distributed. The delta method is used to approximate the variance of the log-*ID* estimate.

The use of the natural log transformation of *ID* estimates to normalize the *ID* distribution is not unusual. In a similar line of this study, Faes et al. (2007) used a log-normal assumption, and their equation (8) is similar to our idea, but with a different algebraic expression. The log transformation of *ID* estimates makes the distribution more symmetric. Even after the log-transformation, skewness still exists. By the same token, Figure 5 in Wheeler and Bailer (2007) showed a skewed distribution at low doses. However, our proposed method with nonparametric quantile estimation further adjusts the skewness via the BCa bootstrap method.

In order to evaluate the actual coverage probabilities of the *ID* LCL estimate using our variance formula accounting for data and model uncertainties, a Monte Carlo simulation study is conducted under linear and nonlinear (sublinear and superlinear) dose-response shapes that can be commonly found in real data sets. A superlinear shape is adopted from *Echovirus* 12 virus data (Teunis et al., 1996) from human volunteers (See Table I), and linear and sublinear shapes are derived with suitable modifications of low-dose

Table I: *Echovirus* 12 virus data (Teunis et al., 1996) from human volunteers

[1]Dose (d_i)	[2]Total (n_i)	[3]Infection (X_i)	[4]Probability (p_i)
330	50	15	0.3000
1000	20	9	0.4500
3300	26	19	0.7308
10000	12	12	1.0000

[1]Dose: ingested numbers of pfu (plaque forming units)
[2]Total: number of subjects at a certain dose
[3]Infection: number of subjects with infection (excretion of echovirus or seroconversion, or both)
[4]Probability: sample proportion (X_i/n_i)

responses. We note that the cited paper (Teunis et al., 1996) contains the *Echovirus* 12 virus data and only linear models and that it is our inference indicating nonlinear response. Under each dose-response shape, one thousand simulation data sets are generated under a binomial assumption. The details are described in Section 3.

2. Methods

2.1 Infectious Dose Estimation

Let I denote the number of independent dose groups. Let n_i denote the number of independent subjects in the i-th group, $i = 1, 2, \ldots, I$. Let X_i denote the number of subjects infected or with symptoms in the i-th group. Assume that X_i has a binomial distribution with n_i and $P(d_i; \boldsymbol{\theta})$, where $P(d_i; \boldsymbol{\theta})$ is a dose-response model and $\boldsymbol{\theta}$ is a parameter vector. The estimation of $\boldsymbol{\theta}$ is accomplished by maximizing the binomial log-likelihood function (Kodell et al., 2002), $l(\boldsymbol{\theta}) \propto \sum_{i=1}^{I}[X_i \ln\{P(d_i, \boldsymbol{\theta})\} + (n_i - X_i)\ln\{1 - P(d_i, \boldsymbol{\theta})\}]$.

In this paper, four two-parameter dose-response models are used as shown in Table II. The two-parameter models include the Beta Poisson (BP), log-normal (LN), log-logistic (LL), and extreme-value (EV) models. The maximization is performed by employing the

Table II: Dose-response models for microbial risk assessment

Name	Model	Domain of Parameters
Beta Poisson (BP)	$P(d; \alpha, \beta) = 1 - \left(1 + \frac{d}{\beta}\right)^{-\alpha}$	$\alpha > 0, \beta > 0$
Log-normal (LN)	$P(d; \alpha, \beta) = \Phi\left(\frac{\ln(d) - \alpha}{\beta}\right)$	$-\infty < \alpha < \infty, \beta > 0$
Log-logistic (LL)	$P(d; \alpha, \beta) = \left[1 + \exp\left(-\frac{\ln d - \alpha}{\beta}\right)\right]^{-1}$	$-\infty < \alpha < \infty, \beta > 0$
Extreme-value (EV)	$P(d; \alpha, \beta) = 1 - \exp[-\exp(\alpha + \beta \ln d)]$	$-\infty < \alpha < \infty, \beta > 0$
Weibull gamma[1] (WG)	$P(d; \alpha, \beta, \gamma) = 1 - \left(1 + \frac{d^\gamma}{\beta}\right)^{-\alpha}$	$\alpha > 0, \beta > 0, \gamma > 0$

[1]Weibull gamma (WG): an additional three-parameter model to establish true underlying models in the simulation study.

SAS procedure NLMIXED (SAS code is available from the authors upon request). It is found that the dose levels may be scaled by a constant scale factor in order to achieve greater stability in the maximum likelihood estimates. The model fitting procedure is invariant to the transformation. Therefore, the estimates of the effective doses are the same regardless of scaling.

When the parameters of a dose-response model are estimated, the effective dose at a specific risk level p (ED_p) for a given model, which itself can be regarded as a fixed "parameter," can be estimated by substituting the "other" parameter estimates and solving for dose d. Every two-parameter model in this study has a closed-form solution for d. If a model has no closed-form solution, the bisection method, a simple and robust root-finding algorithm, can be employed to obtain the estimate of ED_p. Estimates of effective dose at risk levels of 1% (ED_{01}) or 10% (ED_{10}) represent how many microorganisms can produce a 1% or 10% increase in infection or illness, relative to the control response (usually assumed to be zero). These effective dose levels (ED_{01} and ED_{10}) correspond to the lower and upper limits of the risk range (1% and 10%) generally recommended for restricting the calculation of BMDs in chemical risk assessment (correspondingly, *ID*s in microbial risk assessment). For quantal data, an excess risk above background risk of 10% is known as the default benchmark risk (BMR) (Nordberg et al., 2007), which here we term the IDR.

In order to estimate an *ID*, effective doses obtained from the two-parameter models (BP, LN, LL and EV) are averaged using Kullback weights. We let m be the dimension of the parameter vector. The *KIC* is defined as $KIC = -2l(\widehat{\boldsymbol{\theta}}) + 3m$. We define the Kullback weight for the j-th model, denoted by $w_j^{KIC} = \exp(-\Delta_j^{KIC}/2)/\sum_{j=1}^{K}\exp(-\Delta_j^{KIC}/2)$, where $\Delta_j^{KIC} = KIC_j - \min(KIC_1, KIC_2, \ldots, KIC_K)$, the sum of the weights $\sum_{j=1}^{K} w_j^{KIC} = 1$, and K represents the total number of candidate dose-response models. The model with the minimum KIC_j has the most contribution to the *ID* estimate. The point estimate of the *ID* is obtained as

$$\widehat{ID}_p = \sum_{j=1}^{K} w_j^{KIC}\widehat{ED}_{p,j}, \quad (1)$$

where $\widehat{ED}_{p,j}$ represents the effective dose estimate from the j-th model at p (1% or 10%) risk level.

2.2 Lower Confidence Limit on Infectious Dose

In this paper a method to construct a lower confidence limit (LCL) on an ID_p at excess infection rate IDR = p is proposed. We assume that the distribution of *ID* estimates is not symmetric but approximately log-normal (see also Faes et al., 2007). The proposed method incorporates model-averaging with KIC-weights (Moon et al., 2005) and our nonparametric quantile estimator z_α^* via the bias-corrected and accelerated (BCa) bootstrap method (Efron, 1987; Efron and Tibshirani, 1993) in order to adjust for both bias associated with empirical estimates of infectious doses (*ID*) and skewness in the sampling distributions of log-*ID*. The main contribution of the paper is the inclusion of model uncertainty and separation of the model uncertainty from data uncertainty to estimate the lower confidence limit on an *ID*.

It is reasonable to assume that each $\widehat{ED}_k$ has expectation ED_k and that the ED_k s themselves have expectation $ID = \sum_{k=1}^{K} w_k^{KIC} ED_k$ as similar to assumptions of Faes et al. (2007). In order to obtain an LCL on an *ID* at excess risk BMR, we propose the variance of *ID* as follows:

$$Var(ID) = \sum_{k=1}^{K} Var(w_k^{KIC} ED_k) + 2\sum_{j=1}^{K}\sum_{k=j+1}^{K} Cov\left(w_j^{KIC} ED_j, w_k^{KIC} ED_k\right), \quad (2)$$

where the variance term can be decomposed to

$$Var\left(w_k^{KIC} ED_k\right) = Var\left\{E\left(w_k^{KIC} ED_k \middle| M_k\right)\right\} + E\left\{Var\left(w_k^{KIC} ED_k \middle| M_k\right)\right\}, \quad (3)$$

and the covariance term can be obtained as

$$Cov\left(w_j^{KIC} ED_j, w_k^{KIC} ED_k\right) = E\left\{Cov\left[\left(w_j^{KIC} ED_j \middle| M_j\right), \left(w_k^{KIC} ED_k | M_k\right)\right]\right\}, \quad (4)$$

where M_k indicates the model *k*. The proposed formula is based on the law of total variance shown in basic statistics textbooks (Devore, 1991; Burnham and Anderson, 2002). It is also similar to one used in Faes et al. (2007). However, a main difference from the one in Faes et al. (2007) is that the weight vector w_k is treated as a random quantity inside the variance and covariance operators rather than a fixed quantity because even the weights vary from dataset to dataset. We note that there is no covariance among the model means other than the underlying variance itself because *ID* estimates for each model do not co-vary in any defined or measurable way. Hence, we consider only data uncertainty in the covariance term. Therefore, equation (2) with equations (3) and (4) can be rewritten as

$$Var(ID) = \sum_{k=1}^{K} Var\left\{E\left(w_k^{KIC} ED_k \middle| M_k\right)\right\} + \sum_{j=1}^{K}\sum_{k=1}^{K} E\left\{Cov\left[\left(w_j^{KIC} ED_j \middle| M_j\right), \left(w_k^{KIC} ED_k | M_k\right)\right]\right\}. \quad (5)$$

In this framework the first and second terms in equation (5) represent model uncertainty and data uncertainty, respectively. We estimate $Var(ID)$ via bootstrapping by generating *B* bootstrap samples and by estimating w_k^{KIC} and ED_k for each dose-response model in each bootstrap sample.

To construct an LCL on an *ID*, we apply the delta method. We assume that $\ln(ID)$ is approximately normally distributed with mean $\ln(\widehat{ID})$ and variance $Var\left[\ln(\widehat{ID})\right]$. The $Var\left[\ln(\widehat{ID})\right]$ can be approximated by $\widehat{Var(ID)}/\widehat{ID}^2$ using the delta method. Then, an LCL on $\ln(ID)$ can be estimated as

$$\widehat{LCL} = \ln\left(\widehat{ID}\right) + z_\alpha^* \frac{\sqrt{\widehat{Var(ID)}}}{\widehat{ID}}, \quad (6)$$

where $\widehat{Var(ID)}$ is obtained by equation (5). A critical value z_α^* corresponding to the $100(1-\alpha)\%$ confidence level is estimated by

$$z_\alpha^* = \frac{\ln\left(\widehat{ID}\right)_{(L)} - m}{\sqrt{v}}, \quad (7)$$

where $\ln\left(\widehat{ID}\right)_{(L)}$ is the LCL estimate for $\ln(ID)$ from the BCa bootstrap method, and m and v are the mean and variance of log-*ID* estimates from the *B* bootstrap samples, respectively. An LCL on the *ID* is obtained by applying the anti-log transformation $\exp(\widehat{LCL})$.

The estimation of $\ln\left(\widehat{ID}\right)_{(L)}$ is summarized as follows: First, from each bootstrap sample, *B* bootstrap estimates of *ID*'s are obtained. Next, the bias correction factor z_0 is obtained

as $\hat{z}_0 = \Phi^{-1}\left(\#\left[\ln\left(\widehat{ID}_b\right) < \ln\left(\widehat{ID}\right)\right]/B\right)$, where $\ln\left(\widehat{ID}_b\right)$ can be estimated from the B bootstrap samples ($b = 1, 2, \ldots, B$), and $\ln\left(\widehat{ID}\right)$ can be estimated from the original data. Next, the acceleration factor a can be obtained as $\hat{a} = \sum_{b=1}^{B}\left[\overline{\ln(ID)^*} - \ln\left(\widehat{ID}_b\right)\right]^3 / \left[6\left(\sum_{b=1}^{B}\left[\overline{\ln(ID)^*} - \ln\left(\widehat{ID}_b\right)\right]^2\right)^{3/2}\right]$, where $\overline{\ln(ID)^*}$ is the mean of the bootstrap estimates $\ln\left(\widehat{ID}_b\right)$. Finally, we estimate $\ln\left(\widehat{ID}\right)_{(L)}$ by calculating $L = \lfloor B \times \alpha_1 \rfloor$, where $\alpha_1 = \Phi[\hat{z}_0 + (\hat{z}_0 + z_\alpha)/(1 - \hat{a}(\hat{z}_0 + z_\alpha))]$.

This process is computationally intensive. Instead of using one thousand or more bootstrap samples, a smaller bootstrap sample size was determined by a simulation study. The mean differences between z_α^* obtained from a bootstrap sample size of 1000 and z_α^* obtained from bootstrap sample sizes less than 1000 were compared using both a *t*-statistic and a Wilcoxon rank sum statistic as a nonparametric alternative. The bootstrap sample size $B = 700$ was selected based on *p*-value > 0.10.

3. MONTE CARLO SIMULATION STUDIES

In order to examine the coverage of the LCL estimate for *ID*, Monte Carlo simulation studies are conducted. Three distinct monotonic dose-response patterns, sublinear, linear and superlinear are considered, which can be commonly found in real data sets (Teunis, et al., 1996). By way of explanation, when two points on a curve are connected, and the curve is above (below) a straight line connecting the points, we say the curve is superlinear (sublinear). A superlinear pattern is derived from the *Echovirus* 12 data, and linear and sublinear patterns are derived from appropriate modifications. We do not intend to make inferences about *Echovirus* 12. We simply use the *Echovirus* data to ground our simulation study. It illustrates some characteristics of actual microbial dose-response data and gives us a superlinear dose-response pattern as a starting point. All five models (BP, LN, LL, EV and WG) in Table II are fitted to each pattern in Table III to establish "true" underlying dose-response curves. The three-parameter WG model is included along with the two-parameter models because of its historical prominence in MRA. However, only the four two-parameter models are used in model averaging to estimate the *ID* and its LCL.

As shown in Table III, four points in each pattern are considered to generate each dose-response pattern. For example, we consider the superlinear pattern. In order to generate the true probabilities at dose levels 330, 1000, 3300 and 10000 based on Table III, each dose-response model is fitted to the four points to obtain true parameters $\boldsymbol{\theta}$. For generating well-distinguishable three dose-response patterns, six dose-levels are geometrically spaced as d_6 = 10000, d_5 = 5000, d_4 = 2500, d_3 = 1250, d_2 = 625, and d_1 = 312.5. For each dose-level, the "true probability" is determined by $p_i = P(d_i \mid \boldsymbol{\theta})$ for i = 1, 2, …, 6 (as shown in Tables IV and V). Similarly, other patterns can be generated.

Note that true probabilities at the six geometrically spaced dose levels are used when

Table III: Three hypothetical dose response patterns

Response Patterns	d = 330	d = 1000	d = 3300	d = 10000
Superlinear	0.24	0.50	0.76	1.00
Linear	0.04	0.10	0.34	1.00
Sublinear	0.02	0.04	0.08	1.00

Table IV: Generation of Monte Carlo simulation data sets from a true model

Dose Level (i)	Dose (d_i)	Total (n_i)	Response (X_i)	True Probability (p_i)
1	3.125×10^2	10 (or 30)	$Bin(n_1, p_1)$	$p_1 = P_{\text{true}}(d = 3.125 \times 10^2; \boldsymbol{\theta})$
2	6.250×10^2	10 (or 30)	$Bin(n_2, p_2)$	$p_2 = P_{\text{true}}(d = 6.250 \times 10^2; \boldsymbol{\theta})$
3	1.250×10^3	10 (or 30)	$Bin(n_3, p_3)$	$p_3 = P_{\text{true}}(d = 1.250 \times 10^3; \boldsymbol{\theta})$
4	2.500×10^3	10 (or 30)	$Bin(n_4, p_4)$	$p_4 = P_{\text{true}}(d = 2.500 \times 10^3; \boldsymbol{\theta})$
5	5.000×10^3	10 (or 30)	$Bin(n_5, p_5)$	$p_5 = P_{\text{true}}(d = 5.000 \times 10^3; \boldsymbol{\theta})$
6	1.000×10^4	10 (or 30)	$Bin(n_6, p_6)$	$p_6 = P_{\text{true}}(d = 1.000 \times 10^4; \boldsymbol{\theta})$

generating Monte Carlo simulation data sets as shown in Tables IV and V. In Table IV, d_i, n_i, and p_i are fixed as true conditions, whereas the response X_i is randomly generated under the binomial assumption with parameters n_i and p_i. To generate M = 1000 simulation data sets for each designed pattern, each dose-response model is used 200 times as a true model $P_{\text{true}}(d; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\alpha, \beta)$ for the two-parameter models and $\boldsymbol{\theta} = (\alpha, \beta, \gamma)$ for the three-parameter WG model. Two sample sizes of $n_i = 10$ and $n_i = 30$ for i = 1, 2, …, 6 are considered, where the sample size 30 is approximately the average sample size per dose level in the *Echovirus* 12 data. It is common to use the number of dose levels $I \leq 6$ in MRA. After fitting the true configurations, true parameters are used to obtain the ID_{true} from each true dose-response model. In order to calculate the coverage rate for each simulation setting, we observe how many LCL estimates "cover" the ID_{true}. That is, the coverage rate is the proportion of simulations for which the LCL on the ID is less than the ID_{true}. 95% is the nominal coverage rate we are trying to achieve.

As proof of concept, a pilot simulation study has been conducted to verify the coverage

Table V: True hypothetical probabilities at 6 dose levels for Monte Carlo samples.

	Superlinear				
Dose (d_i)	BP	LN	LL	EV	WG
3.125×10^2	0.1884	0.1983	0.1974	0.2202	0.2202
6.250×10^2	0.3347	0.3734	0.3702	0.3540	0.3540
1.250×10^3	0.5407	0.5799	0.5841	0.5359	0.5359
2.500×10^3	0.7611	0.7662	0.7705	0.7404	0.7404
5.000×10^3	0.9171	0.8945	0.8891	0.9065	0.9065
1.000×10^4	0.9822	0.9621	0.9504	0.9845	0.9844
	Linear				
Dose (d_i)	BP	LN	LL	EV	WG
3.125×10^2	0.0605	0.0052	0.0071	0.0110	0.0110
6.250×10^2	0.1174	0.0366	0.0295	0.0350	0.0350
1.250×10^3	0.2209	0.1529	0.1137	0.1083	0.1083
2.500×10^3	0.3931	0.3989	0.3515	0.3084	0.3085
5.000×10^3	0.6316	0.6955	0.6959	0.6948	0.6948
1.000×10^4	0.8643	0.8997	0.9062	0.9781	0.9781
	Sublinear				
Dose (d_i)	BP	LN	LL	EV	WG
3.125×10^2	0.0417	0.0000	0.0001	0.0004	0.0004
6.250×10^2	0.0817	0.0019	0.0010	0.0026	0.0026
1.250×10^3	0.10567	0.0298	0.0101	0.0162	0.0163
2.500×10^3	0.2888	0.1899	0.0964	0.0974	0.0975
5.000×10^3	0.4942	0.5506	0.5268	0.4732	0.4735
1.000×10^4	0.7441	0.8712	0.9208	0.9819	0.9818

of the *ID* via the proposed variance formula when the LN model is known to be the true model. The results from four methods are compared: (1) bootstrap percentile method (Efron and Tibshirani, 1993) without the variance formula (2) the BCa bootstrap method without the variance formula, (3) the variance formula with $z_{0.05} = -1.645$, and (4) the variance formula with estimated critical value using the BCa bootstrap method (denoted by z_α^*). As shown in Table VI, the proposed method with z_α^* outperforms the other three methods, and the coverage rate is close to the nominal level 0.95. We expect that the application of the BCa bootstrap method in the computation of z_α^* should correct bias and skewness if they exist.

After the pilot study using the single LN model, a Monte Carlo simulation study is conducted to evaluate the coverage probabilities obtained from the four methods under the consideration of data uncertainty and model uncertainty. Figures 1 to 3 are obtained by fitting the dose-response patterns in Table III using the four two-parameter models. They serve as the true underlying patterns for sublinear, linear, and superlinear, respectively. Since the BP model appears to be insufficiently flexible to generate a sublinear pattern, it may not be a good representation for the true sublinear pattern (See Figure 1). For only the sublinear pattern, each model is used 250 times as a true model, and two results with and without the BP model are compared. The results from the four methods are compared.

As shown in Table VII, the proposed method also appears to outperform the other methods in model-averaging. The bootstrap percentile method without the variance formula and the BCa bootstrap method without the variance formula do not provide enough coverage, and none of the results reaches the nominal coverage rate 0.95. Using the proposed variance formula, the coverage rate is generally improved and close to 0.95. It seems clear that the proposed variance formula performs better than the bootstrap percentile and the BCa bootstrap methods. Moreover, it is remarkable that using a nonparametric quantile estimate z_α^*, coverage approaches the 0.95 level more tightly than using $z_{0.05} = -1.645$, and this phenomenon is consistent throughout all three configurations regardless of ID_{01} or ID_{10} and $n_i = 10$ or $n_i = 30$. For $n_i = 10$ in the sublinear pattern, none of the four methods has a plausible result. It may be because the pattern is too extreme with a small sample size per dose group.

4. Discussion

Table VI: Results from the pilot simulation study (LN model only)

		[1]Coverage		[2]Coverage		[3]Coverage		**[4]Coverage**	
Pattern	n_i	ID_{01}	ID_{10}	ID_{01}	ID_{10}	ID_{01}	ID_{10}	ID_{01}	$\boldsymbol{ID_{10}}$
Superlinear	10	0.875	0.880	0.933	0.930	0.923	0.904	0.969	**0.954**
	30	0.890	0.889	0.956	0.951	0.928	0.920	0.963	**0.959**
Linear	10	0.832	0.847	0.905	0.898	0.921	0.904	0.953	**0.939**
	30	0.867	0.875	0.932	0.935	0.916	0.921	0.972	**0.956**
Sublinear	10	0.755	0.793	0.853	0.849	0.951	0.913	0.962	**0.916**
	30	0.840	0.841	0.927	0.919	0.896	0.900	0.967	**0.952**

[1]Coverage: coverage rate from bootstrap percentile method
[2]Coverage: coverage rate from bootstrap BCa method
[3]Coverage: coverage rate from the proposed variance formula with $z_{0.05} = -1.645$
[4]Coverage: coverage rate from the proposed variance formula with z_α^*

The main contribution of the paper is the inclusion of model uncertainty, and the separation of model uncertainty from data uncertainty to estimate the lower confidence limit on an *ID*. The proposed method incorporates model-averaging with KIC-weights (Moon et al., 2005) and our nonparametric quantile estimator z_α^* via the BCa bootstrap method in order to adjust for both bias of the *ID* estimate and skewness in the sampling distributions of log-*ID*. We note that the estimation of a lower confidence limit on an *ID* should not be sensitive to the choice of KIC-weights or AIC-weights when the dose-response models have the same number of parameters. In our study, the number of parameters *m* is 2 for every model in model-averaging.

Although the typical BCa bootstrap method (without the proposed variance formula) accounts for both bias and skewness in data uncertainty to improve the bootstrap percentile method, it does not properly account for model uncertainty. As a result, the coverage rates do not meet the desired confidence level 0.95 (as shown in columns 5 and 6 of Tables VI and VII). This result highlights the importance of properly accounting for model uncertainty in a model-averaging method. Our nonparametric quantile estimator z_α^* using the BCa bootstrap method consistently outperforms z_α from the normality assumption regardless of dose-response patterns, risk levels, and sample sizes. In other words, ignoring bias and skewness may lead to inaccurate coverage of the LCL on the *ID*.

In the pilot study in Section 3, the single LN model with the proposed method exhibits nominal coverage when the true setting is designed by the same LN model as shown in Table VI. On the other hand, the result may not be guaranteed for the case when the truth is generated from other models. Wheeler and Bailer (2007) also have shown that using the best model (in terms of χ^2 Goodness-of-Fit statistic) does not provide a satisfactory

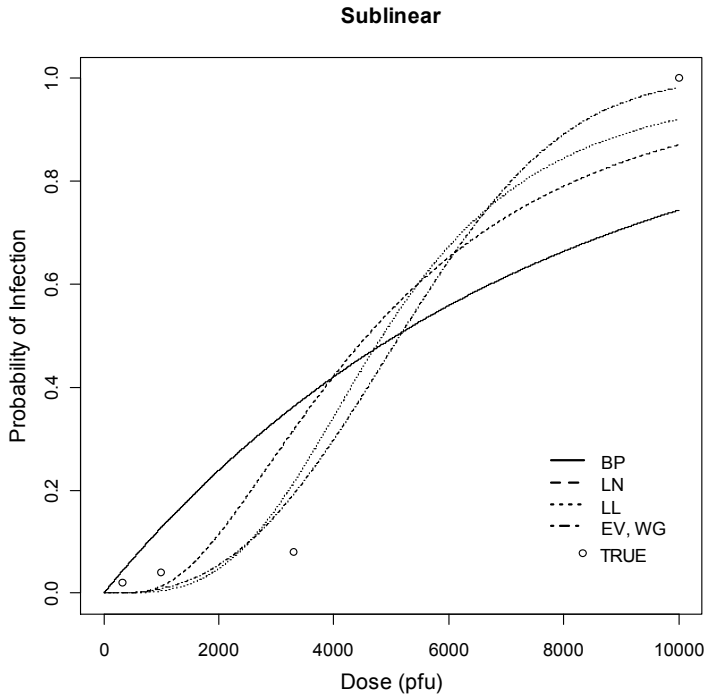

Figure 1. True dose response models fitted to the hypothetical **sublinear** pattern in Table 3. Note that the fits from EV and WG appear to be almost identical.

coverage rate in their simulation study based on various true settings. Since there is no default model in MRA, incorporating model uncertainty along with data uncertainty is recommended to ensure public health protection.

In the Monte Carlo simulation study, the three configurations, superlinear, linear, and sublinear, represent various microbial risk patterns showing the relationship between the dose to an agent and the severity of associated adverse response relating to the food-borne contamination process. The coverage rates from the proposed method in various simulation settings are near the nominal level except for sublinear with $n_i = 10$ as shown in Table VII. It may be due to a small sample size and an extreme sublinear pattern, which may lead to abnormally high estimates of ID_p. None of the four compared methods meets the nominal coverage for the sublinear pattern with the small sample size. For $n_i = 10$ and the given sublinear pattern, it is more likely to have zero binary responses in the low dose-levels, and the estimates of ID_p become abnormal.

The coverage of ID_{01} in the superlinear pattern with $n_i = 30$ appears to be lower than the nominal level (Table VII). It may be due to inadequate representation of model uncertainty because every model is fitted close to each other as shown in Figure 3. The figure indicates that an indistinguishable amount of model variation exists near the 1% level. Elimination of wrong models may be a possible remedy, but the determination of wrong models is another challenge. Furthermore, if a wrong model exists, it has been already treated by the Kullback information criterion by imposing a small or negligible weight in the estimation as a penalty for the poor fit.

Another issue is the sample size for each dose group. In the superlinear pattern, the

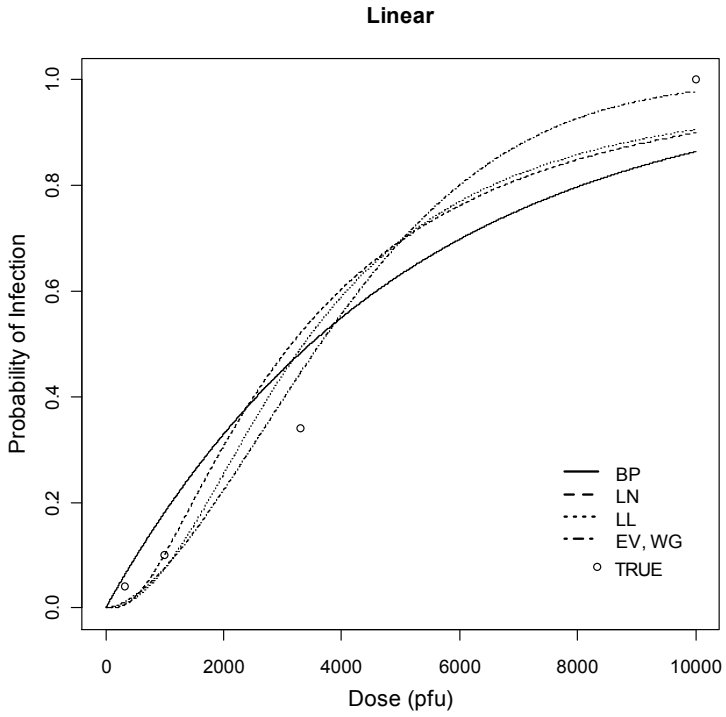

Figure 2. True dose response models fitted to the hypothetical **linear** pattern in Table 3. Note that the fits from EV and WG appear to be almost identical.

Table VII: Coverage rates of the four methods with the four two-parameter models

Pattern	n_i	[1]Coverage		[2]Coverage		[3]Coverage		[4]Coverage	
		ID_{01}	ID_{10}	ID_{01}	ID_{10}	ID_{01}	ID_{10}	ID_{01}	ID_{10}
Superlinear	10	0.6980	0.8230	0.7890	0.8950	0.9020	0.9270	0.9490	0.9670
	30	0.5560	0.7430	0.7110	0.9070	0.8750	0.9150	0.8990	0.9690
Linear	10	0.7020	0.7940	0.8350	0.8780	0.9530	0.9740	0.9520	0.9580
	30	0.7330	0.8230	0.8610	0.9200	0.9390	0.9920	0.9580	0.9810
Sublinear	10	0.6792	0.7888	0.8296	0.8552	0.8360	0.8824	0.8384	0.8880
([5]Including BP)	30	0.6648	0.7504	0.8624	0.9344	0.9688	0.9912	0.9632	0.9752
Sublinear	10	0.7100	0.8050	0.8230	0.8330	0.8190	0.8680	0.8160	0.8760
([6]Excluding BP)	30	0.6840	0.7370	0.8720	0.9310	0.9880	0.9940	0.9720	0.9790

[1]Coverage: coverage rate from bootstrap percentile method
[2]Coverage: coverage rate from bootstrap BCa method
[3]Coverage: coverage rate from the proposed variance formula with $z_{0.05} = -1.645$
[4]Coverage: coverage rate from the proposed variance formula with z_α^*
[5]Including BP: each model is used 250 times as a true model, so $M = 1250$.
[6]Excluding BP: each model is used 250 times as a true model, so $M = 1000$.

coverage for ID_{01} is consistently lower among the methods when the sample size per dose group is higher. In other words, increasing the sample size produces an inaccurate result, which is a contradiction to statistical common sense. Overall, the difference between the sample sizes $n_i = 10$ and $n_i = 30$ seems to be sensitive to patterns and/or risk levels. The exact relationship is still not revealed, and a further investigation is deferred to a future study.

Moon et al. (2004) claimed that the two-parameter models (beta-Poisson, log-normal, log-logistic, and extreme value) reflected at least as much model uncertainty on average as the three-parameter models (Weibull gamma, exponential gamma, Weibull

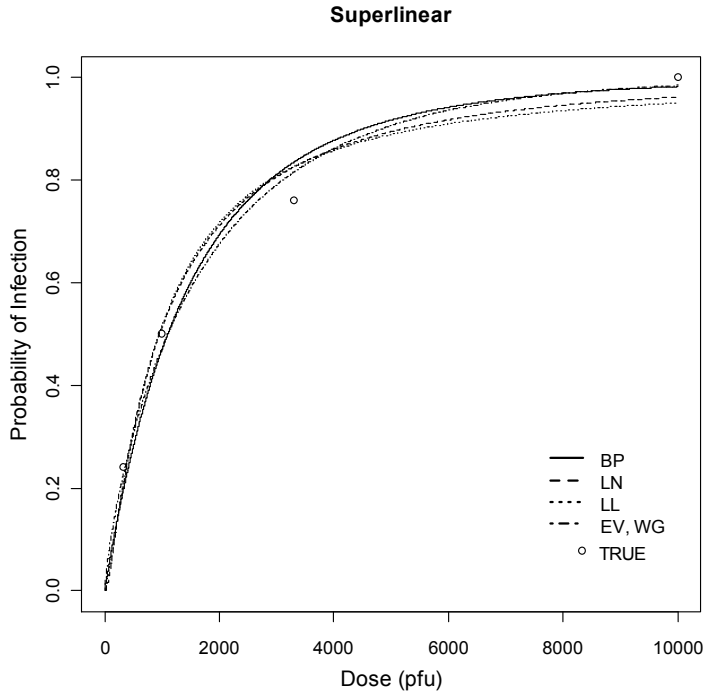

Figure 3. True dose response models fitted to the hypothetical **superlinear** pattern in Table 3. Note that the fits from EV and WG appear to be almost identical.

exponential, and shifted Weibull). Namata et al. (2008) studied model-averaging in MRA with $K = 40$ dose-response models including the same four two-parameter models in this study plus the family of fractional polynomial models with the combination of 3 negative powers and 4 positive powers. They claimed that the set of candidate models should be rich enough. We note that the richness may not be solely determined by the number of dose-response models considered in model-averaging. A set of fewer but diverse models may reflect as much model uncertainty in a more efficient manner. In our simulation study, we confirm that only the four two-parameter models are sufficient to account for model uncertainty in all three representative dose-response patterns.

An ideal combination of dose-response models is unknown, or it may not exist. Our proposed method can be applied with any set of dose-response models. We note that the number of models in model-averaging and the number of model parameters may significantly influence the computational process. We also note that an alternative approach may be model selection based on classification of a model as "mechanistic" or empirical. Mechanistic models are plausible because of their interpretability. However, the underlying assumptions are sometimes strong, and there may be circumstances in which the assumptions are not valid. A good mix of mechanistic and empirical models is also a key point in model-averaging (selection of model space). If data arise from the assumed mechanism, and the mechanistic model fits the data well, then the model will be highly weighted. If assumptions are not met, our estimates will be weighted more by empirical models.

In future studies, we may investigate if a subset of BP, LN, LL, and EV models performs well for all three representative patterns. We may be able to discover a pattern-specific subset of the two-parameter models or of any larger model spaces. For a simple illustration, if a real data set exhibits a sublinear pattern, the BP model appears to be eliminated because it is unable to fit the pattern. In the BP model, the second-derivative with respect to d (dose) is negative for all $\alpha > 0$, $\beta > 0$, and $d > 0$; hence it is unable to be concave upward for any data points (See Figure 1). We note that Teunis et al. (1996) suggest that the BP model is the best-fit model for the *Echovirus* data set, although we assert that the *Echovirus* data exhibit a superlinear dose-response pattern. If a general guideline can be developed for a preferable pattern-specific subset of dose-response models, it may contribute substantially to MRA and it can be widely used in practice.

Acknowledgements

Hojin Moon's research was partially supported by the Research, Scholarship, and Creative Activity (RSCA) Award from CSULB and was partially supported by the Faculty Research Participation Program at the NCTR administered by the Oak Ridge Institute for Science and Education through an interagency agreement between USDOE and USFDA. The views presented in this paper are those of the authors and do not necessarily represent those of the U.S. Food and Drug Administration. Steven Kim's research was originated at CSULB while he was doing his Master's thesis work.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19, 716–723.

Bailer, A. J., Noble, R. B, Wheeler, M. W. 2005. Model uncertainty and risk estimation for experimental studies of quantal responses. Risk Analysis, 25, 291-299.

Burnham, K. P., Anderson, D. R. 2002. Model selection and multimodel inference: a practical information – theoretic approach. 2nd ed. Springer-Verlag.

Cavanaugh, J. E. 1999. A large-sample model selection criterion based on Kulback's symmetric divergence. Statistics and Probability Letters, 42, 333–343.

Centers for Disease Control and Prevention (CDC). 2011. CDC Estimates of Foodborne Illness in the United States. CS218786-A, http://www.cdc.gov/foodborneburden/2011-foodborne-estimates. html, (last visited on 12/31/2011).

Devore, J. L. 1991. Probability and Statistics for Engineers and Scientists, Duxbury Press, Belmont, CA, p.213.

Efron, B. 1987. Better Bootstrap Confidence Intervals. Journal of the American Statistical Association, 82(397), 171–185.

Efron, B., Tibshirani, R. 1993. An Introduction to the Bootstrap, Boca Raton, FL: Chapman & Hall/CRC.

EPA (U.S. Environmental Protection Agency). 2000. *Benchmark Dose Technical Guidance Document* (External Review Draft), EPA/630/R-00/001, October.

Faes, C., Aerts, M., Geys, H., Molenberghs, G. 2007. Model averaging using fractional polynomials to estimate a safe level of exposure. *Risk Analysis*, 27, 111-123.

Farber, J. M., Ross, W. H., Harwig, J. 1996. Health risk assessment of *Listeria monocytogenes* in Canada. International Journal of Food Microbiology, 30, 145-154.

FDA/FSIS (U.S. Food and Drug Administration/Food Safety and Inspection Service). 2003. Quantitative Assessment of Relative Risk to Public Health from Foodborne *Listeria monocytogenes* among Selected Categories of Ready-to-Eat Foods. FDA/CFSAN and USDA/FSIS, September.

Furumoto, W. A. and Mickey, R. 1967. A mathematical model for the infectivity-dilution curve of tobacco mosaic virus: Theoretical considerations. Virology, 32, 216–223.

Please verify that all references are included. Please verify that all references are included. Please verify that all references are included.

Haas, C. N. 1983. Estimation of risk due to low doses of microorganisms: A comparison f alternative methodologies. American Journal of Epidemiology, 118, 573-582.

Haas, C. N., Rose, J. B. and Gerba C. P. 1999. *Quantitative Microbial Risk Assessment.* New York:Wiley.

Hoeting, J. A., Madigan, D., Raftery, A. E, Volinsky, C. T. 1999. Bayesian model averaging: A tutorial. Statistical Science, 14, 382-417.

Kang, S. H., Kodell, R. L., Chen, J. J. 2000. Incorporating model uncertainties along with data uncertainties in microbial risk assessment. Regulatory Toxicology and Pharmacology, 32, 68–72.

Kass, R. E. and Wasserman, L. 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. Journal of the American Statistical Association, 90, 928-934.

Kim, H., Cavanaugh, J. E. 2005. Model section criteria based on Kullback information measures for nonlinear regression framework. Journal of Statistical Planning and Inference, 134, 332-249.

Kodell, R. L., Kang, S, H. and Chen, J. J. 2002. Statistical models of health risk due to microbial contamination of foods. Environmental and Ecological Statistics, 9, 259–271.

Marks, H. M., Coleman, M. E., Lin, C. J. and Roberts, T. 1998. Topics in microbial risk assessment: Dynamic flow tree process. Risk Analysis, 18, 309–328.

Moon, H., Chen, J. J., Gaylor, D. W., Kodell, R. L. 2004. A comparison of microbial dose-response models fitted to human data. Regulatory Toxicology and Pharmacology, 40, 177–184.

Moon, H., Kim, H., Chen, J. J., Kodell, R. L. 2005. Model averaging using the Kullback information criterion in estimating effective doses for microbial infection and illness. Risk Analysis, 25, 1147–1159.

Namata, H., Aerts, M., Faes, C., Teunis, P. 2008. Model averaging in microbial risk assessment using fractional polynomials. Risk Analysis, 28, 891–905.

Nordberg, G. F., Fowler, B. A., Nordberg, M., Friberg, L. 2007. Handbook on the Toxicology of Metals. Section 5.3, Academic Press. Burlington, MA.

Pinsky, P. F. 2000. Assessment of risk from long term exposure to waterborne pathogens. Environmental and Ecological Statistics, 7, 155–175.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics, 6, 461-464.

Teunis, P. F. M., van der Heijden, O. G., van der Giessen, J. W. B., Havelaar. A. H. 1996. The dose-response relation in human volunteers for gastro-intestinal pathogens. Report Number 284550002, National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.

Wheeler, M. W., Bailer, A. J. 2007. Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Analysis, 27, 659–670.

WHO (World Health Organization). *Hazard Identification, Hazard Characterization and Exposure Assessment of* Campylobacter *spp. in Broiler Chickens*. Joint FAO/WHO Activities on Risk Assessment of Microbiological Hazards in Foods-Preliminary Report. Geneva: WHO. 2001a.

WHO (World Health Organization). *Hazard Identification, Exposure Assessment and Hazard Characterization of* Vibrio *spp. in Seafood*. Joint FAO/WHO Activities on Risk Assessment of Microbiological Hazards in Foods – Preliminary Document. Geneva: WHO. 2001b.

WHO (World Health Organization). *Risk Assessments of Salmonella in Eggs and Broiler Chickens*. Microbiological Risk Assessment Series 1. Geneva: WHO. 2002.