A Comparative study of the Estimation of the Maximum Tolerated Dose

Seung Jun Shin * and Sujit K. Ghosh *

Last updated on November 20, 2011

NC State University Department of Statistics Technical Report# 2638

Abstract

Dose-response analysis (DRA) has been a very popular topic within statistical estimation especially in biomedical areas. In addition to estimating the entire dose-response relationship, assessing the risk itself at a certain dose level has been of much interest in some applications, such as in estimating the maximum tolerated dose (MTD) in Phase I clinical trials. Although a variety of parametric dose-response models have been proposed and successfully used in practice, many of the exiting models may not be appropriate in terms of capturing the underlying shape of the dose-response curve. As the estimation of the MTD depends crucially on the assumed dose-response curve, we explore both parametric and nonparametric methods. In addition, we compare both frequentist and Bayesian MTD estimation methods and propose new approaches that are based on Bernstein polynomials to model a dose-response function, which is

*Seung Jun Shin is a graduate student and Sujit K. Ghosh is a Professor, both in the Department of Statistics at North Carolina State University, Raleigh, NC 27695-8203. *Emails:* sshin@ncsu.edu and sujit.ghosh@ncsu.edu

simple yet flexible enough to capture various shape restriction. One of the main benefits of employing the Bernstein dose-response model is that the statistical inference for the MTD estimator is as straightforward as the parametric models since it can be shown that the MTD estimation reduces to a constraint linear model. It is shown via empirical studies that the proposed MTD estimation methods outperform the popular log-logit model under the violation of logistic assumptions and are competitive and robust against parametric model misspecifications.

Keywords: *Bernstein polynomial; dose-response model; maximum tolerated dose; toxicity probability function.*

1 Introduction

Modeling dose-response relationships has been extensively studied over the past several decades especially in biomedical applications such as toxicology, epidemiology and clinical trials. Dose-response analysis (DRA) attempts to measure the change in effect on living organisms like human or animals (response) caused by different levels (dose) of a stressor. The response can be recorded either on a continuous scale or as categorical, however quantal (or dichotomous) response is used most commonly in practice. With quantal responses, the dose-response relationship can be summarized by an underlying probability function of dose, the so called toxicity probability function.

We illustrate the methodologies by using data from a toxicological study conducted at the National Center for Toxicological Research(NCTR). The study examined the effect of exposure to the different levels of herbicide 2,4,5-trichlorophenoxyacetic acid. One outbred

(CD-1) and four inbred (C57BL/6, C3H/He, BALB/C, and A/JAX) strains of mice were investigated. Further details of the study and design are given in Holson et al. (1992) and raw data for A/JAX mice are available in Ahn and Chen (1997). The original data contain seven dose levels (0, 15, 20, 25, 30, 45, and 60 mg/kg/day) of trichlorophenoxyacetic aid but we adjust response at dose level 0 to impose restrictions described in Section 2. For simpler illustration we collapse the raw data by dose and Figure 1 shows the empirical toxicity probabilities based on the collapsed dose-response data. Clearly the dose-response curve does not appear to reassemble a logistic curve. We present both parametric and nonparametric analysis of this data in Section 5.

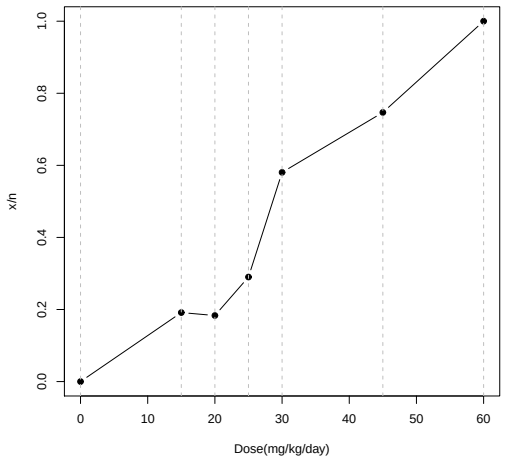


Figure 1: Empirical toxicity probability function of malfunctioned mice(x_j) out of administered ones(n_j) at different dose level, d_j ($j = 0, 1, \cdots, 6$) from the toxicology study conducted at the National Center for Toxicology Research (Holson et al., 1992).

In practice, there are two main objectives in DRA. The first is to test whether there exist any changes in effect at different exposure levels. Usual ANOVA-type approaches would be appropriate for this purpose. However it is often of interest to assess the risk at a certain level of exposure as a secondary objective of DRA and ANOVA-type approaches may not

be effective for this task as the risk can not be assessed at an un-measured dose level. For example, in Phase I clinical trials one of the main goals is to determine the maximum dose of a new pharmacological treatment with significant effect but that will not cause unacceptable adverse effect; commonly known as the maximum tolerated dose (MTD). Along this purpose, regression-type approaches have become more popular than ANOVA-type in DRA. A majority of commonly used regression approaches impose a parametric model to the toxicity probability function and then parameters are estimated using maximum likelihood approach. Generalized linear models (GLM), including logit and probit models are the most popular ones in DRA and they have been routinely used for many decades. Threshold models have also been used in toxicology (Cox, 1977). Armitage-Doll multistage model (Peto, 1977) and gamma multihit model (Cornfield, 1977; Rai and Van Ryzin, 1979) have received considerable attentions as well. An extensive list of currently available parametric dose-response models are enumerated in Bailer, Noble, and Wheeler (2005).

Despite the availability of a wide range of dose-response models discussed above, it is not clear which model is appropriate in practice. The parametric dose-response models presume certain shapes of underlying toxicity probability functions in advance and hence they perform poorly if such assumptions are violated. Indeed, it is well known that the maximum likelihood estimator (MLE) is inconsistent under the missspecification of the parametric form of the likelihood function. For instance, logit model can not produce a non-sigmoidal probability curve (e.g., see Figure 1 or (b) and (c) in Figure 2.) and hence the corresponding MTD estimate will be biased with smaller variance than what it should be. A natural remedy to this drawback is to consider a class of more flexible models. Bailer et al. (2005) proposed the use of the Bayesian model averaging (BMA) approach which weighs each parametric

model and uses posterior model probabilities to obtain a single model averaged across a finite number of models using BIC type approximation of Bayes factors. However, such a class of models though very flexible is still restricted by the class of parametric models considered by the researcher.

As an alternative to the rich class of parametric models, one may also consider nonparametric models. Common nonparametric function estimation methods can be readily extended to estimate the toxicity probability functions by imposing some constraints on the functional shape of the curve. However it is a common trade-off that the nonparametric model makes the subsequent statistical inferences much more complicating and may lose efficiency of the estimates. For example, it is not straightforward to obtain the sampling distribution of a MTD estimator based on a nonparametric dose-response model whereas such estimates can be readily obtained in the logit model by applying the standard large sample theories.

In this article, we propose both of frequentist and Bayesian MTD estimation methods based on Bernstein polynomials to estimate the dose-response relationship by extending the idea of Curtis and Ghosh (2011). They proposed to use Bernstein polynomials to approximate an unknown function in the regression problem with the monotonicity constraint and also connected the Bernstein-polynomial-based model to a standard linear model by considering a simple reparametrization. We extend their idea to a single dose-response model, which will subsequently be referred to as the Bernstein dose-response model. The Bernstein dose-response model is not only flexible enough to capture a wide variety of potential shape of the curve but also provides a straightforward method to conduct further statistical inference by adopting common linear model methodologies.

The remainder of the paper is organized as follows. Various dose-response models including both classical parametric models and the proposed Bernstein dose-response model are introduced in Section 2. Both frequentist and Bayesian MTD estimation methods are described in Section 3. We carry out numerical studies in Section 4 to compare the performance of the proposed MTD estimators based on the nonparametric Bernstein dose-response model to the existing methods. The proposed MTD estimators are illustrated in Section 5 by applying them to the real data obtained from NCTR. Finally, conclusions with remarks follow in Section 6.

2 Dose-Response Models

Dose-response models with quantal exposure can be characterized by a binomial distribution at different dose levels as follow. Let

$$Y_j \sim Bin(n_j, \pi_j), \qquad j = 1, \cdots, J$$

where Y_j is the number of affected subjects out of a total of n_j subjects administered at a dose d_j and J is the number of dose levels considered. Let $\pi_j = \pi(d_j)$ denote the true toxicity probability at dose level d_j. Estimating the entire toxicity probability function $\pi(\cdot)$ based on the data $\{(Y_j, n_j, d_j); j = 1, \cdots, J\}$ is the primary interest in DRA. In practice, the function $\pi(\cdot)$ is assumed to be continuous and nondecreasing which satisfies $0 = \pi(0) \leq \pi(d) \leq \pi(\infty) = 1$ for any dose $d \geq 0$.

2.1 Parametric Models

Parametric dose-response models are based on specifying a continuous and nondecreasing function that is allowed to depend on a finite dimensional parameter $\boldsymbol{\theta} \in \theta \subset \mathbb{R}^k$ for some finite $k \geq 1$. Since most of the restrictions on the $\pi(\cdot)$ function are the same as those of a continuous cumulative distribution function (CDF), it is natural to consider the following family of parametric models.

$$\pi_\theta(d) = \begin{cases} F(\log d; \boldsymbol{\theta}) & \text{if support}(F) = \mathbb{R} \\ F(d; \boldsymbol{\theta}) & \text{if support}(F) = [0, \infty) \end{cases} \tag{1}$$

where F denotes a known continuous CDF such as logistic, normal, gamma or Weibull distributions. The most popular choice of F is the logistic distribution function, known as the log-logit model whose toxicity probability function depends on $\boldsymbol{\theta} = (\theta_0, \theta_1)^T$ and is given by

$$\pi_\theta(d) = \left\{1 + e^{-\theta_0 - \theta_1 \log d}\right\}^{-1}, \tag{2}$$

where $\theta_0 \in \mathbb{R}$ and $\theta_1 \geq 0$ are parameters to be estimated. Parametric models are simple in the sense that they are characterized only through a finite number of parameters (e.g., $k = 2$ in case of log-logistic model) and many useful theoretical results are readily obtained by employing standard large sample theories based on the likelihood function. Maximum likelihood approach is the most common method to estimate $\boldsymbol{\theta}$ and hence to estimate π_θ by plugging the MLE of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$ in the toxicity probability function, i.e., $\hat{\pi}(d) = F(\log d; \hat{\boldsymbol{\theta}})$. Bayesian approaches can also be used by introducing prior distributions of $\boldsymbol{\theta}$ and obtaining correspond-

ing posterior distributions of $\boldsymbol{\theta}|Data$. The Bayesian estimator of π_θ is given by computing an appropriate posterior summary. For instance, we can obtain $\hat{\pi}(d) = E\{F(\log d; \boldsymbol{\theta})|Data\}$ where the expectation is taken with respect to the posterior distribution of $\boldsymbol{\theta}$ given the data $\{(Y_j, n_j, d_j) : j = 1, \cdots, J\}$.

2.2 Nonparametric Models

One of the common drawbacks for the parametric models is the lack of flexibility which often leads to biased and inconsistent estimates. It is well known that if the true function $\pi(\cdot)$ does not belong to the assumed parametric family which is often the case in real applications, the MLE of $\boldsymbol{\theta}$ is inconsistent and hence the MLE of $\pi_\theta(\cdot)$ would also be invalid. Nonparametric dose-response models are not based on a particular family of F for $\pi_\theta(\cdot)$. Within the statistics and econometric literature, nonparametric regression with shape restrictions has been extensively studied for a long time. The kernel and the spline techniques are the two of the most popular nonparametric approaches and many related novel methods have been developed since the 1970s (Ramsay, 1988; Mukerjee, 1988; Mammen et al., 2001; Hall and Huang, 2001; Wang and Li, 2008). On the other hand, Bayesian approaches have also appeared on similar topics (Brown and Chen, 1999; Perron and Mengersen, 2001; Bornkamp and Ickstadt, 2009; Curtis and Ghosh, 2011). Stadtmüller (1986) was one of the first to introduce Bernstein polynomials in the context of nonparametric function estimation. Since then, related studies have been popular especially for nonparametric Bayesian methods (Tenbusch, 1997; Brown and Chen, 1999). Curtis and Ghosh (2011) proposed Bayesian nonparametric estimation for regression functions with shape restrictions via Bernstein polynomials. In

this article, we extend their approach to estimate the toxicity probability function $\pi(\cdot)$ in dose-response model.

Bernstein polynomials (Lorentz, 1986) of degree M for a continuous function $\pi(x)$ where $x \in [0,1]$ is defined as follows:

$$\mathbb{B}_M(\pi(x)) = \sum_{k=0}^{M} \pi\left(k/M\right) \binom{M}{k} x^k (1-x)^{M-k} = \sum_{k=0}^{M} \beta_k b_M(x,k) \equiv B_M(x;\boldsymbol{\beta}), \tag{3}$$

where $\beta_k = \pi\left(k/M\right)$; $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_M)^T$; and $b_M(x,k)$ is the probability mass of the Binomial random variable at $k = 0, 1, \cdots, M$ with parameters M and x. Weierstrass approximation theorem states that $\mathbb{B}_M(\pi(\cdot))$ uniformly converges to $\pi(\cdot)$ as $M \to \infty$. Therefore the nonparametric model of the toxicity probability function, $\pi(d)$ based on a sequence of Bernstein polynomials as M increases can be considered as follows:

$$\pi_M(d) \;\; = \begin{cases} \sum_{k=0}^{M} \beta_k b_M\left(\frac{d}{D}, k\right) & \text{if } 0 \le d \le D \\ \beta_M + (1-\beta_M)\frac{d-D}{d-D+1} & \text{if } d > D \\ \text{subject to } 0 = \beta_0 \le \cdots \le \beta_M \le 1, \end{cases} \tag{4}$$

where $\beta_k = \pi\left(\frac{kD}{M}\right)$. It is not difficult to convince ourselves that the model (4) satisfies all the required constraints of a toxicity probability function: (i) $\pi_M(\cdot)$ is continuous and nondecreasing and (ii) $\pi_M(0) = 0 \le \pi_M(d) \le \pi_M(\infty) = 1, \forall d > 0$. Thus the estimation of $\pi(\cdot)$ reduces to the estimation of a finite dimensional parameter $\boldsymbol{\beta}$ where M can be chosen as function of J, e.g., $M = o(J^b)$ for some $b \in (0.2, 0.4)$ (Stadtmüller, 1986; Tenbusch, 1997). We note that the lower equation in (4) depends only through β_M and hence we can focus only on the upper equation. The value $D = \inf\{d > 0 : \pi(d) = 1\}$ can be

estimated by the maximum dose administered in the study. Curtis and Ghosh (2011) link the model (4) to the standard linear model by employing a simple reparametrization given by $\gamma_k = \beta_k - \beta_{k-1}, k = 1, \cdots, M$ and $\beta_0 = \gamma_0 = 0$, thus $\beta_k = \sum_{i=1}^{k} \gamma_i$. Now, we can re-express Bernstein polynomials (3) in terms of γ_ks $(k = 1, \cdots, M)$ as follows.

$$B_M(x, \beta) = \sum_{k=1}^{M} \beta_k b_M(x, k) = \sum_{k=1}^{M} \left(\sum_{l=1}^{k} \gamma_l \right) b_M(x, k) = \sum_{l=1}^{M} \gamma_l F_M(x, l),$$

where $F_M(x, l) = \sum_{k=l}^{M} b_M(x, k)$. Moreover it is easy to show that

$$F_M(x, l) = \sum_{k=l}^{M} b_M(x, k) = \int_0^x \frac{u^{l-1}(1-u)^{M-l}}{\mathrm{B}(l, M-l+1)} du,$$

where $\mathrm{B}(l, M-l+1)$ denotes the Beta function and the last integral denotes the distribution function of the Beta random variable with parameters l and $M - l + 1$. Consequently, we have a nonparametric model of $\pi(\cdot)$ in terms of $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_M)^T$ (and M) for $0 \leq d \leq D$ as follows.

$$\begin{aligned} &\pi_M(d) = \textstyle\sum_{k=1}^{M} \gamma_k F_M\left(\frac{d}{D}, k\right) = \mathbf{F}_M^T(d)\boldsymbol{\gamma} \\ &\text{subject to } \begin{pmatrix} \mathbf{I}_M \\ -\mathbf{1}_M^T \end{pmatrix} \boldsymbol{\gamma} \geq \begin{pmatrix} \mathbf{0}_M \\ -1 \end{pmatrix} \end{aligned} \tag{5}$$

where $\mathbf{F}_M(d) = \left(F_M\left(\frac{d}{D}, 1\right), \cdots, F_M\left(\frac{d}{D}, M\right)\right)^T$, and $\mathbf{1}_M, \mathbf{0}_M$ and $\mathbf{I}_M$ are M dimensional vectors of ones, zeros and identity matrix, respectively. Notice that the model (5) turns out to be linear in $\boldsymbol{\gamma}$ and hence a linear-model-based methodology such as the least square approach can be applied to estimate $\boldsymbol{\gamma}$. The degree of Bernstein polynomials, M can be

regarded as a tuning parameter of the model and popular model selection methods including the cross-validation can be applied to select an optimal M.

3 MTD Estimation Methods

The MTD is generally defined as the largest dose of a pharmacological treatment that will produce the desired effect without unacceptable toxicity. Estimating the MTD is one of the main goals in DRA. The MTD at a desired tolerance level $\alpha \in (0, 1)$ which controls acceptable toxicity, denoted by $d^*(\alpha)$ is defined as follows:

$$d^*(\alpha) = \sup\{d > 0 : \pi(d) \leq \alpha\}. \tag{6}$$

In practice, it is common to use $\alpha = 0.33$ but other values can also be used.

3.1 Parametric Estimation

As the parametric models (1) always produce continuous and strictly increasing functions of π_θ, the definition of the MTD (6) is reduced to $d^*(\alpha) = \pi_\theta^{-1}(\alpha)$ where π_θ^{-1} is the inverse function of π_θ. For example, the log-logit model (2) provides a closed form of the MTD as

$$d^*_{Flogit}(\alpha) = \pi_\theta^{-1}(\alpha) = e^{-\frac{\theta_0}{\theta_1}} \left(\frac{1}{\alpha} - 1\right)^{-\frac{1}{\theta_1}} \equiv g_\alpha(\boldsymbol{\theta}).$$

By the invariance property, the MLE of the MTD, $\hat{d}^*_{Flogit}(\alpha)$ is obtained by $g_\alpha(\hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ denotes the MLE of $\boldsymbol{\theta}$ which maximizes the log-likelihood function given by

$$l(\boldsymbol{\theta}) = \sum_{j=1}^{J} \left[y_j \log \pi_\theta(d_j) + (n_j - y_j) \log \left\{ 1 - \pi_\theta(d_j) \right\} \right],$$

where $\pi_\theta(d)$ is defined in (2). Applying the *delta method*, an asymptotic distribution of $\hat{d}^*_{Flogit}(\alpha) = g_\alpha(\hat{\boldsymbol{\theta}})$ is given by

$$\sqrt{n}(g_\alpha(\hat{\boldsymbol{\theta}}) - g_\alpha(\boldsymbol{\theta})) \sim AN\left(0, \dot{g}_\alpha(\boldsymbol{\theta})\{I(\boldsymbol{\theta})\}^{-1}\{\dot{g}_\alpha(\boldsymbol{\theta})\}^T\right), \tag{7}$$

where $\dot{g}_\alpha(\boldsymbol{\theta}) = \frac{\partial g_\alpha(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$, $I(\boldsymbol{\theta})$ is Fisher information matrix and AN denotes 'asymptotically normally distributed'. The corresponding interval estimator of the MTD can be directly constructed from the distribution (7) by plugging the MLE in the expression of the asymptotic variance in (7).

Bayesian MTD estimator of a parametric model is developed by introducing priors on $\boldsymbol{\theta}$. For example, if we consider again the log-logit model, the prior distributions can be specified as

$$\theta_0 \sim N(0, \sigma_0^2) \qquad \text{and} \qquad \theta_1 \sim \text{Uniform}(a_0, b_0), \text{ where } 0 \leq a_0 < b_0. \tag{8}$$

The restrictions on $\pi(\cdot)$ function suggest using large σ_0^2 and nonnegative a_0, b_0 in practice. Posterior inference about $\boldsymbol{\theta}$ and hence about $\pi_\theta(d)$ can be obtained numerically via Markov Chain Monte Carlo (MCMC) method and the corresponding Bayesian MTD estimator from log-logit model, denoted by $d^*_{Blogit}(\alpha)$ is the posterior expectation, $E(g_\alpha(\boldsymbol{\theta})|Data)$. An inter-

val estimator can be constructed by computing the highest posterior density region employing the algorithm proposed by Hyndman (1996) after generating a sufficient number of posterior samples using the MCMC method. We implemented the MCMC method using the popular `WinBUGS` software.

3.2 Nonparametric Estimation

The Bernstein polynomial model (5) produces a strictly increasing probability function if $\gamma_k > 0$ for some k and the MTD estimator is determined by estimating $\boldsymbol{\gamma}$ and a properly selected M. We consider both frequentist and Bayesian approaches to obtain estimators of $\boldsymbol{\gamma}$ as well as the MTD.

3.2.1 Approximate Frequentist Method

First, we consider a nonparametric frequentist method based on the large sample theories. For large n_j, the empirical toxicity probability $\hat{\pi}_j = \frac{y_j}{n_j}$ follows

$$\sqrt{n_j}(\hat{\pi}_j - \pi_j) \sim AN\left(0, \pi_j(1-\pi_j)\right).$$

by Central Limit Theorem. When $Y_j = 0$ for some j, we use the Anscombe correction for $\hat{\pi}_j = \frac{Y_j+3/8}{n_j+3/4}$. Now, $\hat{\pi}_j$ can be regarded as a response variable in a linear regression and $\boldsymbol{\gamma}$ is

estimated by the weighted least square (WLS) method as follows.

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \sum_{j=1}^{J} \hat{w}_j(\hat{\pi}_j - \pi_j)^2 = \arg\min_{\boldsymbol{\gamma}} \sum_{j=1}^{J} \hat{w}_j(\hat{\pi}_j - \mathbf{F}_M(d_j)^T\boldsymbol{\gamma})^2 \tag{9}$$

$$\text{subject to} \quad i) \quad \gamma_k \geq 0, k = 1, \cdots, M \quad \text{and} \quad ii) \quad \sum_{k=1}^{M} \gamma_k \leq 1.$$

where $\hat{w}_j = \frac{n_j}{\hat{\pi}_j(1-\hat{\pi}_j)}$. The optimization problem (9) can be readily solved via quadratic programming (e.g., using `quadprog` in `R`), and hence we can obtain the nonparametric estimate of the MTD, say the approximate frequentist MTD estimator $\hat{d}^*_{NPF}(\alpha)$, by solving the nonlinear equation $\hat{\pi}_M(d) = \mathbf{F}_M^T(d)\hat{\boldsymbol{\gamma}} = \alpha$ (in terms of d). Any numerical root-finding algorithm (e.g, `uniroot` in `R`) can be adopted to obtain $\hat{d}^*_{NPF}(\alpha)$. As $\hat{\pi}_M(d)$ is strictly increasing in d, such a root will be unique. It is not straightforward to derive an asymptotic distribution of $\hat{d}^*_{NPF}(\alpha)$, and we propose to use a resampling method to obtain a confidence interval pre-pivoted via double bootstrapping (Shao and Tu, 1995).

Alternatively, we can also use a variance stabilizing transformation for $\hat{\pi}_j$ by noting the following:

$$\sqrt{n_j}\left(\sin^{-1}\sqrt{\hat{\pi}_j} - \sin^{-1}\sqrt{\pi_j}\right) \sim AN\left(0, \frac{1}{4}\right).$$

We note that $\sin^{-1}\sqrt{\pi(d)}$ is an increasing function of d bounded by $\pi/2$ and hence $\sin^{-1}\sqrt{\pi(d)}$ itself can be approximated by Bernstein polynomials of degree M as

$$\sin^{-1}\sqrt{\pi_M(d)} = \mu_M(d) = \mathbf{F}_M^T(d)\boldsymbol{\eta}$$

The estimator of $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_M)^T$ is obtained again by WLS method as follows:

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \sum_{j=1}^{J} 4n_j(\hat{\mu}_j - \mu_j)^2 = \arg\min_{\boldsymbol{\gamma}} \sum_{j=1}^{J} 4n_j(\hat{\mu}_j - \mathbf{F}_M(d_j)^T\boldsymbol{\eta})^2$$
$$\text{subject to} \quad i) \quad \eta_k \geq 0, k = 1, \cdots, \quad \text{and} \quad ii) \quad \sum_{k=1}^{M} \eta_k \leq \pi/2,$$

where $\hat{\mu}_j = \sin^{-1}\sqrt{\hat{\pi}_j}$. Both point and interval estimators of the MTD can be obtained from $\hat{\boldsymbol{\eta}}$ in a similar way to described above. One of the benefits of adopting variance stabilizing transformation is that we can eliminate nuisance quantities, w_j in the objective function and it may improve the performance of the MTD estimator.

Selection of degrees of Bernstein polynomial, M is another important issue in Bernstein nonparametric model. We propose to use leave-one-out cross-validation based on minimizing the corresponding WLS metric. Note that M cannot be greater than J in the frequentist approach as it would make the design matrix within our WLS framework singular.

3.2.2 Bayesian Nonparametric Method

Next we develop a nonparametric Bayesian MTD estimation for the Bernstein dose-response model, which provides finite-sample estimators. Clearly we can not use arbitrary priors for $\boldsymbol{\gamma}$ directly because it may not satisfy the constraints, $\gamma_k \geq 0, k = 1, \cdots, M$ and $\sum_{k=1}^{M} \gamma_k \leq 1$. Instead, we consider the following reparametrization, similar to a stick-breaking construction:

$$\gamma_1 = \delta_1 \qquad \text{and} \qquad \gamma_k = \delta_k \prod_{i=1}^{k-1}(1 - \delta_i), \quad k = 2, \cdots, M.$$

It is easy to show that $\sum_{k=1}^{M} \gamma_k = 1 - \prod_{k=1}^{M}(1-\delta_k)$ and hence $\boldsymbol{\gamma}$ satisfies the required constraints as long as the new parameters satisfy $0 \leq \delta_k \leq 1$ for all $k = 1, \cdots, M$. As δ_ks are unrestricted on the range $[0, 1]$, independent Beta distributions can be used as a prior, i.e, $\delta_k \overset{\text{iid}}{\sim} Beta(a_\delta, b_\delta)$ can be a natural choice. Alternatively, we also consider the following prior employing the idea proposed by Curtis and Ghosh (2011):

$$P(\delta_k) = 0.5 \cdot \mathbf{1}(\delta_k = 0) + 0.5 \cdot \mathbf{1}(0 < \delta_k < 1), \tag{10}$$

which allows a point mass on 0 with probability 0.5. We often observe in dose-response analysis that the toxicity probability curve is nearly flat over certain regions. In order to capture such nearly flat regions, we propose to use the prior given in (10). Note that if $\delta_k = 0$ and consequently $\gamma_k = 0$ then $\beta_{k-1} = \beta_k$ represents a nearly flat region for $d \in \left[\frac{(k-1)D}{M}, \frac{kD}{M}\right]$ especially when M is large. This motivates the use of point mass on $\delta_k = 0$ in (10). By employing similar sampling scheme of Curtis and Ghosh (2011), we can obtain posterior sample of $\boldsymbol{\gamma}$ via MCMC algorithm and compute corresponding posterior samples of MTDs by solving $\mathbf{F}_M^T(d)\boldsymbol{\gamma} = \alpha$ for each of the sampled $\boldsymbol{\gamma}$s using a root-finding method. Finally the nonparametric Bayesian MTD estimator based on the Bernstein dose-response model, denoted by $\hat{d}^*_{NPB}$ is the posterior mean of the sampled roots of the equations. An interval estimator can be constructed by the highest posterior density region method as we did in the parametric Bayesian model. We can choose an optimal M by employing the Bayesian model selection method based on the deviance information criterion (DIC) as follows. We start with $M = J$ and compute DIC as well as the effective number of parameters. Then we use the effective number of parameters as a new M and continue to update M until it

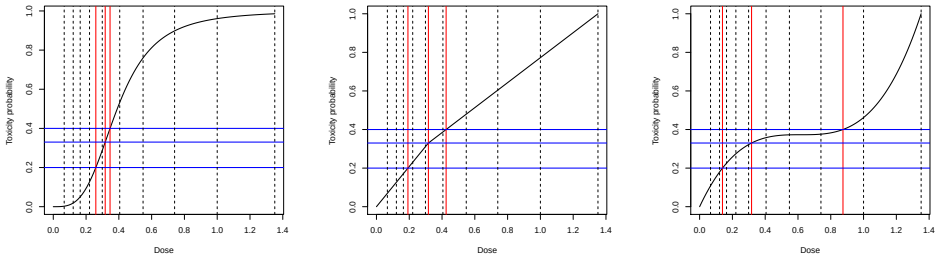
stabilizes.

4 Simulation Study

We carry out Monte Carlo investigation to compare the performances of different MTD estimators discussed in Section 3. To generate simulated data, we fix the dose levels, d_js and the corresponding number of subjects, n_j as in the real data reported by Thompson and Funderlic (1981). They recorded the number of mice affected by benzpyrene injection with ten different dose levels, $J = 10$. In order to evaluate the performance of each MTD estimator, we consider the following three different shapes of toxicity probability functions. Note that the non-logit functions π_2 and π_3 plateaus at 1 beyond $d \geq 1.351$.

- Logit - $\pi_1(d) = \{1 + e^{-3.2-3.4 \log d}\}^{-1}$
- Piecewise Linear - $\pi_2(d) = \begin{cases} 1.04d & \text{if } 0 \leq d \leq 0.317 \\ 0.125 + 0.648d & \text{if } 0.317 < d \leq 1.351 \\ 1 & \text{if } d > 1.351 \end{cases}$
- Cubic - $\pi_3(d) = \begin{cases} 1.586(d - 0.617)^3 + 0.373 & \text{if } 0 \leq d \leq 1.351 \\ 1 & \text{if } d > 1.351 \end{cases}$

The choice of the true parameter $\boldsymbol{\theta} = (3.2, 3.4)^T$ of logit function π_1 was taken from the MLE of Thompson and Funderlic's data under log-logit model. Note that the parameters of π_2 and π_3 are chosen such that $\pi_1^{-1}(0.33) = \pi_2^{-1}(0.33) = \pi_3^{-1}(0.33)$. Figure 2 shows the true toxicity probability functions and the corresponding true MTDs at the three different levels of $\alpha = 0.2, 0.33$ and 0.4. The vertical dashed lines represent the 10 dose levels used for the study.



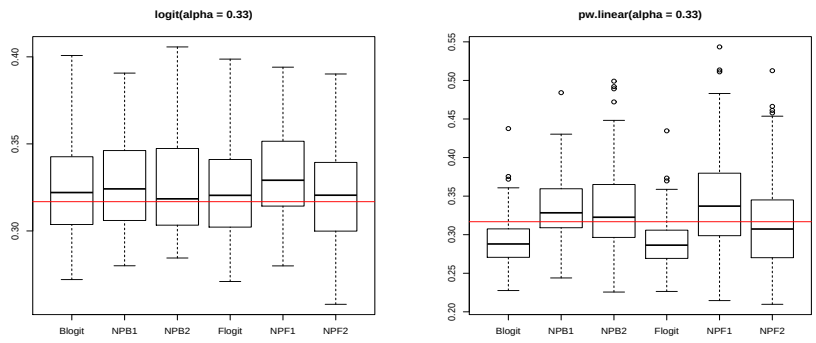
(a) π_1 - Logit (b) π_2 - Piecewise Linear (c) π_3 - Cubic

Figure 2: Three types of toxicity probability functions, π_1, π_2 and π_3 and corresponding true MTDs (red vertical lines) at different $\alpha = 0.2, 0.33, 0.4$ (blue horizontal lines). Note that all the true MTDs at $\alpha = 0.33$ in the three functions are identical. The dashed vertical lines represent the dose levels used to simulate data.

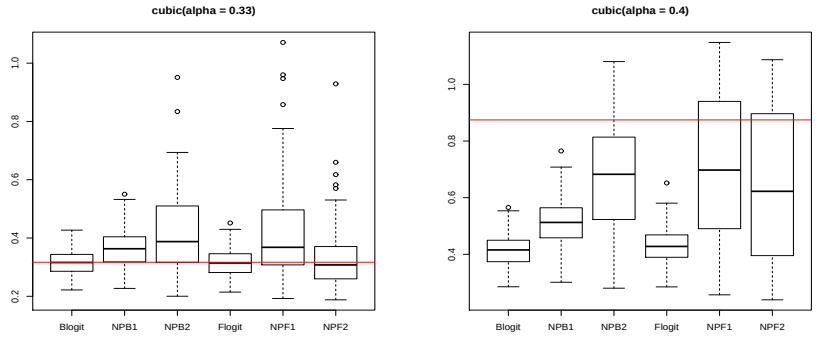
We compare 6 different MTD estimators which include both frequentist and Bayesian MTD estimators, $\hat{d}^*_{Flogit}(\alpha)$ and $\hat{d}^*_{Blogit}(\alpha)$ based on the log-logit model described in Section 3.1; two approximate frequentist nonparametric MTD estimators based on the two different asymptotic distributions, $\hat{d}^*_{NPF1}(\alpha)$ and $\hat{d}^*_{NPF2}(\alpha)$ described in Section 3.2.1; Bayesian nonparametric MTD estimators with Beta and zero mass prior (10), respectively denoted by $\hat{d}^*_{NPB1}(\alpha)$, $\hat{d}^*_{NPB2}(\alpha)$ as described in Section 3.2.2.

For $\hat{d}^*_{Blogit}(\alpha)$, normal and uniform priors (8) are used with $\sigma_0^2 = 100$, $a_0 = 0$ and $b_0 = 10$. For $\hat{d}^*_{NPB1}(\alpha)$, we set the hyperparameters, $a_\delta = b_\delta = 1$ of Beta distribution. In order to obtain approximate samples from the posterior distribution for Bayesian methods, we implemented a MCMC algorithm with 11,000 iterations, discarded the first 1,000 iterations as burn-in and used a final chain with 10,000 samples for posterior inference.

We evaluate the bias and the root mean squared error (RMSE) of our estimators based on 100 Monte Carlo replications summarized in Figure 3 and Table 1. It is noteworthy that we observe all four nonparametric estimators are quite competitive when the true function



(a) MTD estimates for $\pi_1^{-1}(0.33)$ (b) MTD estimates for $\pi_2^{-1}(0.33)$



(c) MTD estimates for $\pi_3^{-1}(0.33)$ (d) MTD estimates for $\pi_3^{-1}(0.4)$

Figure 3: The figures show boxplots of MTD estimates in selected MC simulation results based on 100 independent repetitions. Parametric estimators are biased under model misspecification while nonparametric estimators are not or at least much less biased. In the cubic function $\pi_3(d)$ case, parametric MTD estimators performs well at $\alpha = .33$, however it is not universal for arbitrary $\alpha \in (0, 1)$. Note that they are severely biased for $\alpha = 0.4$ in (d).

is $\pi_1(d)$, although the parametric ones, $\hat{d}^*_{Flogit}(\alpha)$ and $\hat{d}^*_{Blogit}(\alpha)$ are overall winners in terms of bias as well as RMSE as expected. On the other hand, nonparametric estimators are robust in terms of having reduced bias against model misspecifications. We note that the nonparametric methods do not use any of the parametric assumptions and that they take into account the potential variability generated from model uncertainty across a wide variety of monotonic curves. As a result, the proposed nonparametric methods tend to have larger variances but much less biases than the existing parametric methods. The severe biases

of the parametric models in estimating MTDs are clearly evident when the true toxicity probability function is not logistic.

We also compare the performances of the corresponding interval estimators. For each of the nonparametric frequentist estimators, we compute two different types of bootstrap confidence intervals, that use both with and without prepivoting via double bootstrap. The eight different interval estimates whose target coverage is 95% are compared in terms of nominal coverage probability and average length, which are presented in Table 2. The parametric estimators are, as we expect, best under the logit case, $\pi_1(d)$, and their performances deteriorate in other cases in terms of not achieving the targeted coverage probability. Under a misspecified model, the parametric methods produce shorter but severely biased intervals. This matches with the rationale that the parametric models fail to take into account model uncertainty and produce biased results under model misspecifications. It is remarkable that most of the nonparametric methods attain a nominal coverage probability of 0.95 even if they are much longer than the parametric ones. Nonparametric methods essentially employ information only from data itself and hence they may show moderate performance in terms of efficiency with regard to small sample but they are asymptotically unbiased. We also note that bootstrap intervals prepivoted via double bootstrapping based on nonparametric frequentist estimators are usually longer compared to the ones from parametric models but they attain 95% nominal coverage. $\hat{d}^*_{NPB2}$ outperforms other nonparametric estimators especially under cubic function, π_3 which has nearly flat region near the middle part of the curve.

We find that $\hat{d}^*_{NPF2}$ performs reasonably well in terms of achieving low bias and RMSE. On the other hand $\hat{d}^*_{NPB2}$ performs quite well in terms of attaining the 95% nominal coverage

$\pi(d)$	α		Blogit	NPB1	NPB2	FLogit	NPF1	NPF2
Logit(π_1)	.2	Bias	.0076	-.0139	-.0069	.0053	.0023	-.0057
			(.0023)	(.0020)	(.0025)	(.0022)	(.0026)	(.0027)
		RMSE	.0239	.0246	.0258	.0230	.0257	.0276
			(.0402)	(.0255)	(.0340)	(.0383)	(.0319)	(.0318)
	.33	Bias	.0067	.0090	.0092	.0050	.0172	.0024
			(.0024)	(.0025)	(.0029)	(.0024)	(.0029)	(.0029)
		RMSE	.0250	.0264	.0303	.0244	.0334	.0293
			(.0365)	(.0334)	(.0505)	(.0353)	(.0443)	(.0358)
	.4	Bias	.0061	.0202	.0174	.0049	.0244	.0064
			(.0025)	(.0027)	(.0031)	(.0025)	(.0030)	(.0031)
		RMSE	.0261	.0336	.0353	.0257	.0386	.0313
			(.0349)	(.0429)	(.0557)	(.0340)	(.0482)	(.0384)
PwLinear(π_2)	.2	Bias	-.0035	.0053	-.0009	-.0053	.0085	-.0077
			(.0028)	(.0030)	(.0030)	(.0028)	(.0040)	(.0036)
		RMSE	.0279	.0304	.0299	.0280	.0410	.0370
			(.0549)	(.0595)	(.0514)	(.0528)	(.0686)	(.0404)
	.33	Bias	-.0257	.0162	.0166	-.0272	.0272	-.0012
			(.0033)	(.0041)	(.0057)	(.0033)	(.0066)	(.0064)
		RMSE	.0419	.0441	.0587	.0426	.0714	.0638
			(.0546)	(.0843)	(.1084)	(.0537)	(.1228)	(.0943)
	.4	Bias	-.0712	-.0144	-.0060	-.0722	.0068	-.0083
			(.0038)	(.0048)	(.0071)	(.0037)	(.0085)	(.0090)
		RMSE	.0804	.0497	.0709	.0812	.0846	.0900
			(.0598)	(.0732)	(.1074)	(.0596)	(.1492)	(.1488)
Cubic(π_3)	.2	Bias	.0352	.0294	.0066	.0189	.0143	.0217
			(.0026)	(.0031)	(.0032)	(.0031)	(.0039)	(.0026)
		RMSE	.0438	.0426	.0321	.0359	.0416	.0340
			(.0530)	(.0657)	(.0614)	(.0556)	(.0951)	(.0655)
	.33	Bias	.0004	.0482	.1007	-.0014	.1056	.0160
			(.0042)	(.0063)	(.0140)	(.0046)	(.0177)	(.0113)
		RMSE	.0420	.0790	.1721	.0455	.2052	.1133
			(.0613)	(.1141)	(.3291)	(.0691)	(.4506)	(.3620)
	.4	Bias	-.4605	-.3649	-.2065	-.4440	-.1620	-.2270
			(.0054)	(.0082)	(.0173)	(.0062)	(.0247)	(.0256)
		RMSE	.4636	.3739	.2687	.4482	.2944	.3409
			(.1062)	(.1608)	(.3056)	(.1211)	(.3289)	(.3502)

Table 1: MC averaged biases and root mean squared errors (RMSE) of MTD estimates based on 100 independent repetitions: Parametric methods are the best the in the log-logit case, $\pi_1(d)$ while they show poor performance in the other cases, π_2 and π_3. Nonparametric methods may produce larger RMSE than parametric ones but they are almost always much less biased under the both π_2 and π_3. The corresponding MC standard errors are in parentheses.

$\pi(d)$	α		Blogit	NPB1	NPB2	FLogit	NPF1		NPF2	
							naive	pivoted	naive	pivoted
Logit(π_1)	.2	Coverage	0.96	0.92	0.94	0.90	0.78	0.90	0.88	1.00
		Length	.0850	.0907	.0828	.0692	.0933	.1480	.1042	.1766
			(.0006)	(.0008)	(.0021)	(.0005)	(.0019)	(.0035)	(.0018)	(.0033)
	.33	Coverage	0.95	0.98	0.95	0.94	0.80	0.96	0.90	1.00
		Length	.0889	.1020	.0945	.0842	.1064	.1670	.1180	.2010
			(.0007)	(.0009)	(.0020)	(.0007)	(.0021)	(.0039)	(.0021)	(.0037)
	.4	Coverage	0.95	0.95	0.94	0.96	0.82	0.99	0.92	1.00
		Length	.0934	.1085	.1009	.0920	.1121	.1756	.1262	.2147
			(.0008)	(.0009)	(.0020)	(.0008)	(.0022)	(.0040)	(.0022)	(.0040)
PwLinear(π_2)	.2	Coverage	0.89	0.94	0.95	0.85	0.90	0.96	0.76	0.85
		Length	.0935	.1240	.1375	.0816	.1977	.3453	.1585	.2618
			(.0010)	(.0019)	(.0040)	(.0011)	(.0059)	(.0078)	(.0041)	(.0069)
	.33	Coverage	0.86	0.97	0.94	0.86	0.89	0.96	0.86	0.92
		Length	.1232	.1787	.2274	.1273	.2634	.4142	.2788	.4708
			(.0014)	(.0028)	(.0065)	(.0018)	(.0073)	(.0095)	(.0079)	(.0095)
	.4	Coverage	0.49	0.94	0.93	0.50	0.81	0.93	0.81	0.90
		Length	.1484	.2061	.2719	.1555	.2956	.4527	.3417	.5470
			(.0020)	(.0032)	(.0069)	(.0024)	(.0082)	(.0111)	(.0092)	(.0100)
Cubic(π_3)	.2	Coverage	0.80	0.94	0.94	0.92	0.88	0.94	0.98	1.00
		Length	.0990	.1299	.1292	.1160	.2861	.5899	.1458	.2895
			(.0014)	(.0028)	(.0042)	(.0025)	(.0155)	(.0238)	(.0055)	(.0144)
	.33	Coverage	0.94	0.95	0.97	0.95	0.78	0.86	0.78	0.85
		Length	.1685	.2885	.5282	.2351	.6505	.8717	.5903	.8144
			(.0025)	(.0061)	(.0161)	(.0062)	(.0174)	(.0130)	(.0200)	(.0140)
	.4	Coverage	0.00	0.03	0.75	0.01	0.50	0.56	0.47	0.52
		Length	.2224	.3582	.6012	.3235	.6790	.8749	.7203	.8804
			(.0034)	(.0055)	(.0110)	(.0095)	(.0128)	(.0090)	(.0108)	(.0039)

Table 2: Coverage probabilities and averaged lengths of eight MTD interval estimators with 95% target coverage based on 100 independent recitations: In the log-logit case; $\pi_1(d)$, the parametric methods outperform the others however, all nonparametric methods (except NPFs without prepivoting) are still competitive. Other than in the ideal log-logit case, parametric interval estimators break down in terms of coverage probability. They are restricted on the logit assumption and hence the variability of the corresponding MTDs are under-estimated, which results in short intervals but are biased under the violation of model assumptions. The corresponding MC standard errors are in parentheses.

with the shortest expected length of the interval. So we recommend the use of $\hat{d}^*_{NPF2}$ and $\hat{d}^*_{NPB2}$ for most practical purposes.

5 Real Data Analysis

We illustrate all the methods with the toxicological study data of 2,4,5-trichlorophenoxyacetic acid (Holson et al., 1992) introduced in Section 1. We also consider another popular dose-response model for real data analysis, the so called Emax model (MacDougall, 2006). Non-linear least square estimation is employed to fit the Emax model. The corresponding MTD estimators are obtained by usual root finding algorithm from the fitted curve of the Emax model and the bootstrap method is used again to compute the corresponding confidence interval. The Emax model is another popular parametric model for the sigmoidal shapes of the true toxicity probability functions but more flexible than the classical logistic and probit models. Further details about the Emax model are given in (MacDougall (2006)). As we impose a restriction, $\pi(0) = 0$, the 0 dose level is excluded and hence we have $J = 6$ dose levels. For the Bayesian models, we ran three independent chains of 11,000 iterations discarded 1,000 as burn-in for each chain and thus finally used the remaining 30,000 MCMC samples for posterior inference. We employ the same priors as the ones used in our simulation study. For the bootstrap confidence intervals based on $\hat{d}^*_{NPF1}(\alpha)$ and $\hat{d}^*_{NPF2}(\alpha)$, we generate 500 bootstrap samples and 100 double bootstrap samples for each of the 500 bootstrapped ones. We found $\hat{M}_{NPF1} = 6$ and $\hat{M}_{NPF2} = 6$ by using leave-one-out cross-validation method for frequentist approaches and $\hat{M}_{NPB1} = \hat{M}_{NPB2} = 6$ from DIC-based method for Bayesian approaches. Figure 4 shows the estimated probability curves based on each model we have

considered. The estimated MTDs at $\alpha = 0.2, 0.33, 0.4$ as well as the corresponding interval estimator are summarized in Figure 5. Bayesian nonparametric estimates, $\hat{d}^*_{NPB1}(\alpha)$ and $\hat{d}^*_{NPB2}(\alpha)$ are significantly different from the others, especially at $\alpha = 0.2$. The differences get statistically insignificant as α increases. Note also that although the Emax model also produces a sigmoidal curve, the corresponding MTD estimates are quite different from the ones obtained under the log-logit model.

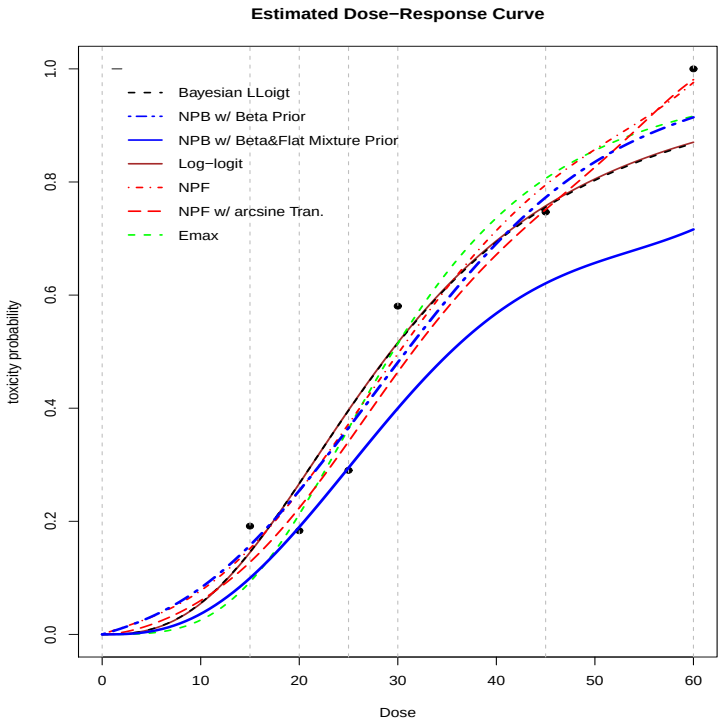
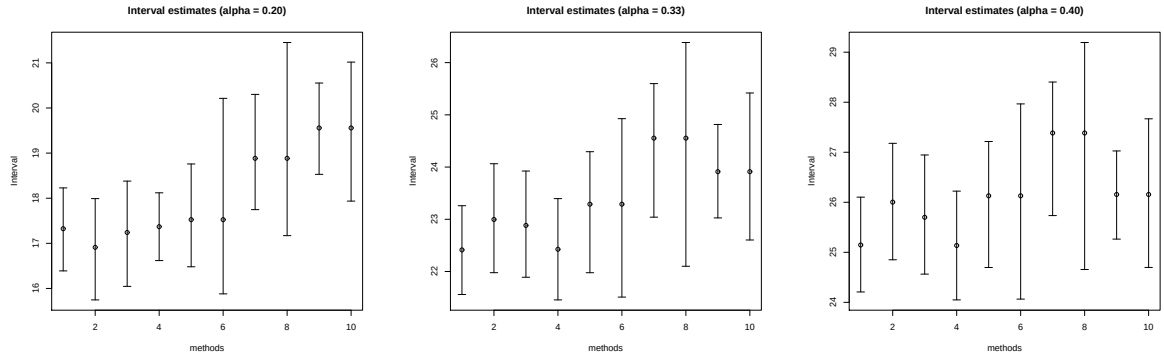


Figure 4: The seven different estimated toxicity probability functions from different methods discussed: Nonparametric models as well as Emax model produce quite different estimated probability curves from each other whereas the two log-logit models produce nearly identical ones.

α	Flogit	Blogit	NPB1	NPB2	NPF1	NPF2	Emax
0.2	17.37 (.383)	17.26 (.463)	13.04 (.317)	18.88 (.646)	17.42 (.770)	17.26 (.618)	19.56(0.537)
0.33	22.43 (.495)	22.40 (.439)	21.51 (.524)	24.55 (.658)	23.65 (.512)	22.88 (.490)	23.91(0.467)
0.4	25.14 (.555)	25.15 (.489)	26.07 (.635)	27.38 (.715)	26.13 (.500)	25.69 (.585)	26.16(0.458)

Table 3: Estimated MTD estimates for Ahn and Chen (1997) data : The MTD estimates for the different levels $\alpha = 0.2, 0.33$ and 0.4 are displayed with the corresponding standard errors in parentheses.



(a) $\alpha = .2$ (b) $\alpha = .33$ (c) $\alpha = .4$

Figure 5: The ten different interval estimates (vertical lines) as well as the seven point estimates (dots in the middle) for Ahn and Chen (1997) data at different $\alpha = 0.2, 0.33, 0.4$: NPF1, NPF2, and Emax models produce two different types of bootstrap interval estimates based on a single point estimate, respectively. The horizontal axis represents indices of the methods; 1:Flogit, 2:Blogit, 3:NPB1(beta), 4:NPB2(zero mass prior), 5:NPF11(naive), 6:NPF12(pivoted); 7:NPF21 (arcsine&naive), 8:NPF22(arcsine&pivoted), 9:Emax(naive), 10:Emax(pivoted).

6 Conclusion

In this paper, we have proposed new MTD estimation methods based on the nonparametric dose-response model using Bernstein polynomials. The proposed model can be represented as a standard linear model framework which enables the adoption of standard linear model approaches for estimation of the true probability function. The new MTD estimators are equipped with benefits for both parametric and nonparametric aspects in the sense that those are not only relatively straightforward to fit as in a parametric linear model approaches but also are flexible enough to capture potential variabilities from model uncertainty. Through extensive simulation studies, we have demonstrated the robustness of our proposed nonparametric methods against model misspecification as well as their competitiveness to parametric models under correct model assumptions. All of the numerical methods presented in this manuscript are programmed using `R` language and are available for the authors upon request.

References

Ahn, H. and Chen, J.J (1997). Tree-structured logistic model for over-dispersed binomial data with application to modeling devlopmental effects. *Biometrics* **53,** 435–455.

Bailer, A.J. Noble, R.B. and Wheeler, M.W. (2005). Model Uncertainty and risk estimation for experimental studies of Qunatal responses. *Risk Analysis* **25,** 291–299.

Bornkamp, B. and Ickstadt, K (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose-reponse analysis. *Biometrics* **65,** 198–205.

Brown, B.M. and Chen, S.X (1999). Beta-Bernstein smoothing for regression curves with compact support. *Scandinavian Journal of Statistics* **26,** 47–59.

Cornfield, J. (1977). Carsenogenic risk assessment. *Scinece* **198,** 693–699.

Cox, C. (1987). Threshold dose-response models in toxicology. *Biometrics* **43,** 511–523.

Curtis, S.M. and S.K. Ghosh (2011). A Variable Selection Approach to Monotonic Regression with Bernstein Polynomials. *Journal of Applied Statistics* (in press).

Hall, P. and Huang, L. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics* **29,** 624–647.

Holson, J.F., Gaines, T.B., Nelson, C.J., LaBorde, J.B., Gaylor, D.W., Sheehan, D.M, et al. (1992). Developmental toxicity of 2,4,5-trichlorophenoxyacetic acid I: Multireplicated dose-response studies in four inbred strains and one outbred stock of mice. *Fundamental and Applied Toxicology* **19,** 286–297.

Hyndman, R.J. (1996). Computing and graphing highest density regions. *American Statistician* **50,** 120–126.

Lorentz, G.G. (1986). *Bernstein Polynomials.* American Mathematical Society.

MacDougall, J. (2006). Analysis of dose-response studies - Emax model, in N. Ting (ed.) *Dose Finding in Drug Development*, New York: Springer

Mammen, E., Marron, J.S., Turlach, B.A., and Wand, M.P. (2001) A general projection framework for constrain smoothing. *Statistical Science*, **16,** 232–248.

Mukerjee, H. (1988). Monotone nonparametric regression. *Annals of Statistics* **16,** 741–750.

Perron, F. and Mengersen, K. (2001). Bayesian nonparametric modeling using mixture of triangular distributions. *Biometrics* **57,** 518–528.

Peto, R (1977). Epidemiology, multistage models and shor-term mutagenicity tests, In Hiatt, H.H., Watson, J.D., and Winston, J.A. (Eds.) *Origin of Human Cancer Book C. Human Risk Assessment.* 1403–1428, New York: Cold Spring Harbor.

Ramsay, J.O. (1988). Monotone regression splines in action. *Statistical Science* **3,** 425-461.

Rai, K. and Van Ryzin, J. (1979). Risk assessment of toxic enviromental substences based on a genralized muti-hit model. In Breslow, N. and Whittmore, A (Eds.) *Energy and Health.* 99–117, Philadelphia: Society of Industrial and Applied Mathematics Press.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* New York: Springer.

Stadtmüller, U. (1986). Asymptotic properties of nonparametric curve estimates. *Periodica Mathematica Hungarica* **17,** 83–108.

Tenbusch, A. (1997). Nonparametric curve estimation with Bernstein estimates. *Metrika* **45,** 1–30.

Thompson, W.A. and Funderlic, R.E. (1981). A simple threshold model for the classical bioassay problem, G.G. Berg and H.D. Maillie (eds) In *Measurement of Risks.* 521–553, New York: Pleunum.

Wang, X. and Li, F. (2008). Isotonic smoothing spline regression. *Journal of Computational and Graphical Statistics* **17,** 21–37