

# A Bayesian Nonparametric Modeling Framework for Developmental Toxicity Studies

Kassandra Fronczyk and Athanasios Kottas \*

**Abstract:** We develop a Bayesian nonparametric mixture modeling framework for replicated count responses in dose-response settings. We explore this methodology for modeling and risk assessment in developmental toxicity studies, where the primary objective is to determine the relationship between the level of exposure to a toxic chemical and the probability of a physiological or biochemical response, or death. Data from these experiments typically involve features that can not be captured by standard parametric approaches. To provide flexibility in the functional form of both the response distribution and the probability of positive response, the proposed mixture model is built from a dependent Dirichlet process prior, with the dependence of the mixing distributions governed by the dose level. The methodology is tested with a simulation study, which involves also comparison with semiparametric Bayesian approaches to highlight the practical utility of the dependent Dirichlet process nonparametric mixture model. Further illustration is provided through the analysis of data from two developmental toxicity studies.

**KEY WORDS:** Dependent Dirichlet process; Developmental toxicology data; Dirichlet process mixture models; Gaussian process; Markov chain Monte Carlo; Risk assessment.

---

\*K. Fronczyk is Postdoctoral Researcher, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 (E-mail: [kmfronczyk@mdanderson.org](mailto:kmfronczyk@mdanderson.org)), and A. Kottas is Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064. (E-mail: [thanos@ams.ucsc.edu](mailto:thanos@ams.ucsc.edu)). The authors wish to thank Marc Mangel for helpful discussions, as well as the Editor, an Associate Editor, and two referees for comments that greatly improved the presentation of the material in the paper. This research is part of the Ph.D. dissertation of Kassandra Fronczyk, completed at University of California, Santa Cruz, and was supported in part by the National Science Foundation under award DEB 0727543, and by a Special Research Grant awarded by the Committee on Research, University of California, Santa Cruz.

# 1 Introduction

## 1.1 Background

Birth defects induced by toxic chemicals are investigated through developmental toxicity studies. In these studies, at each experimental dose level, a number of pregnant laboratory animals (dams) are exposed to the toxin and the number of resorptions (i.e., undeveloped embryos or early fetal deaths) and/or prenatal deaths, the number of live pups, and the number of live malformed pups from each dam are typically recorded. Additional outcomes measured on each of the live pups may include body weight and length.

The main purpose of developmental toxicity studies is to examine the relationship between the level of exposure to the toxin (dose level) and the probability of malformation (or, in general, response). The dose-response curve is defined by the probability of an outcome across the dose levels. Also of interest is quantitative risk assessment, which evaluates the probability that adverse effects may occur as a result of the exposure to the substance. While the objectives of the studies are clear, the resulting data are a veritable gold mine of statistical challenges. Many of these difficulties arise from the inherent heterogeneity in the data due to the clustering of individuals within a group and the variability of the reaction of the individuals to the toxin. Another challenging feature of the data is associated with the multiple related outcomes, both continuous (e.g., body weight) and discrete (e.g., number of malformations).

A variety of approaches for the analysis of developmental toxicity studies have been suggested in the statistical literature. Modeling approaches based on standard parametric response distributions and/or customary parametric forms for dose-response curves include Chen et al. (1991), Catalano and Ryan (1992), Ryan (1992), Zhu et al. (1994), and Regan and Catalano (1999). However, due to the various sources of heterogeneity, data from many studies indicate vast departures from parametric models. A different line of research has focused on classical semiparametric or likelihood estimation for the joint distribution of the vector of binary responses associated with each dam under the assumption of exchangeability (e.g., Bowman and George,

1995; George and Bowman, 1995; Kuk, 2004; Pang and Kuk, 2005). Although such approaches provide more general modeling for the response distribution than traditional parametric models, dose-response relationships are still introduced through parametric forms. Moreover, inferential challenges include interpolation at unobserved dose levels (a key objective for risk assessment) as well as uncertainty quantification for point estimates.

By comparison to likelihood and classical semiparametric approaches, Bayesian methods have not been widely used for the analysis of developmental toxicity studies. Examples of parametric Bayesian hierarchical models for toxicology data, comprising joint discrete-continuous outcomes, include Dunson et al. (2003) and Faes et al. (2006). To our knowledge, the only Bayesian semiparametric model is presented by Dominici and Parmigiani (2001), using a product of mixtures of Dirichlet process prior structure.

## 1.2 Data examples

In a Segment II developmental toxicity experiment,  $n_i$  pregnant dams are exposed to dose level,  $x_i$ ,  $i = 1, \dots, N$ . Dam  $j = 1, \dots, n_i$  at dose  $x_i$  has  $m_{ij}$  implants, of which the number of resorptions ( $r_{ij}$ ) and prenatal deaths ( $d_{ij}$ ) are typically recorded as  $R_{ij} = r_{ij} + d_{ij}$ , and the number of live pups at birth with a certain defect are recorded as  $y'_{ij}$ . Consequently, the litter size (the number of viable fetuses) for dam  $j$  at dose  $x_i$  is  $m_{ij} - R_{ij}$ . The outcomes from the  $j$ th dam at dose level  $x_i$  may be recorded as  $\{(m_{ij}, y_{ij}) : i = 1, \dots, N, j = 1, \dots, n_i\}$ , where  $y_{ij} = R_{ij} + y'_{ij}$ , or, more generally, as  $\{(m_{ij}, R_{ij}, y'_{ij}) : i = 1, \dots, N, j = 1, \dots, n_i\}$ . The prevailing data structure found in the statistical literature appears to be of the first type, where the random variables involved are the number of implants and the sum of all negative outcomes.

Thus, each triplet of data  $(m, y, x)$ , comprising the number of implants, number of negative outcomes, and dose level, defines a particular dam. Two data sets commonly considered in the statistical literature for developmental toxicity experiments are shown in Figure 1. The left panel plots a data set from a toxicity study regarding the effects of the herbicide 2,4,5-trichlorophenoxyacetic (2,4,5-T) acid (Holson et al., 1991). We work with the version of the data

given in Table 3 of Bowman and George (1995), where the number of combined endpoints consists of the number of resorptions and prenatal deaths, and the number of fetuses with cleft palate malformation. The experiment considers  $N = 6$  doses, one control and 5 active dose groups. The number of animals per dose level ranges from 25 to 97 dams. The number of implants ranges from 1 to 21 across all dams and all dose levels, with 25th, 50th and 75th percentiles given by 10, 12 and 13, respectively. Based on exploratory analysis, the data suggest varying departures from the Binomial model across the dose levels, indicating the need for a flexible model to capture the evolution of the response distributions over the range of dose levels.

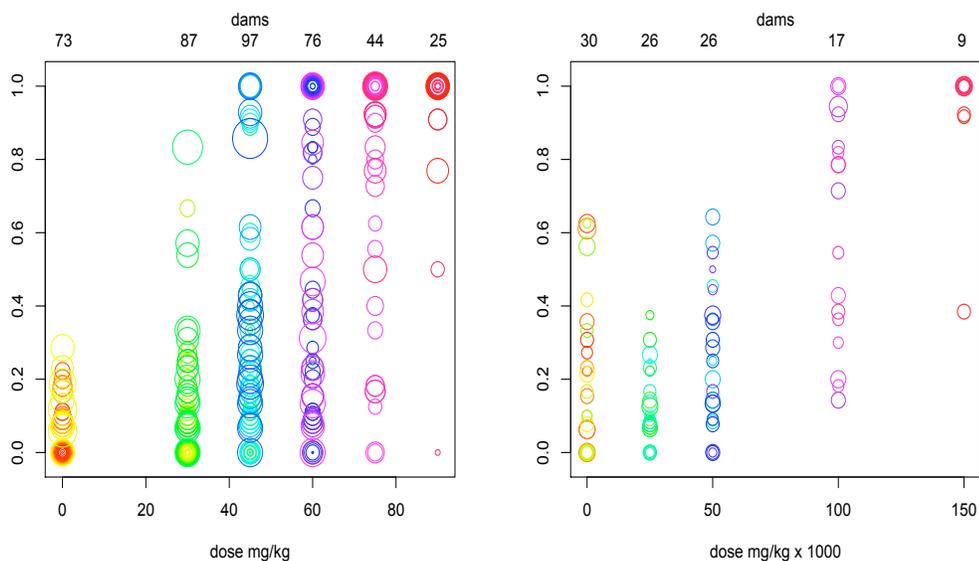


Figure 1: Plots of the 2,4,5-T data (left panel) and the DEHP data (right panel). Each circle corresponds to a particular dam, the size of the circle is proportional to the number of implants, and the coordinates of the circle are the dose level and the proportion of combined negative outcomes. Shown at the top of each panel is the number of animals per dose level. The online version of this figure is in color.

The second data set (Figure 1, right panel) is from an experiment that explored the effects of diethylhexalpthalate (DEHP), a commonly used plasticizing agent. It is known that these plasticizers may leak in small quantities from plastic containers with various solvents such as food or milk. The possibility of toxic effects from these agents have been recognized and tested in

developmental toxicity studies such as the one described in Tyl et al. (1983). The DEHP study is also discussed by Molenberghs and Ryan (1999), although they consider a different version of the data set than the one available from the database of the National Toxicology Program (which is the version we work with). Here, the combined endpoints include resorption, prenatal death, and malformation of a live fetus (external, visceral or skeletal malformation).

The number of dams per dose level is about a third of those found in the 2,4,5-T data; the number of implants across all dams and dose levels ranges from 4 to 18, with 25th, 50th, and 75th percentiles equal to 11, 13, and 14, respectively. Particularly noteworthy is the drop in the proportions of combined negative outcomes from dose 0 to 25 mg/kg  $\times$ 1000, which may indicate a hormetic dose-response relationship. Hormesis refers to a dose-response phenomenon characterized by favorable biological responses to low exposures to toxins, and thus by opposite effects in small and large doses. For endpoints involving disease incidence (e.g., mutation, birth defects, cancer), hormesis results in a J-shaped dose-response curve. Although the possibility of different low dose effects is accepted, the suggestion of positive low dose effect is debated, hence, hormesis is a controversial concept in the toxicological sciences (e.g., Calabrese, 2005). Notwithstanding the ultimate scientific conclusions, to be able to uncover non-standard dose-response relationships, we need a modeling framework for the dose-dependent response distributions which enables flexible inference for the implied, possibly non-monotonic, dose-response curve.

### **1.3 Objectives and outline**

To overcome the limitations of parametric approaches, and at the same time retain a fully inferential framework, we develop a Bayesian nonparametric mixture model that provides flexibility in both the response distribution and the dose-response relationship. We seek mixture modeling for response distributions that are related across doses with the level of dependence driven by the distance between the dose values. To this end, we consider a dependent Dirichlet process (DDP) prior for the dose-dependent mixing distributions. Inference and prediction under DDP prior structures requires replication, which arises through the number of dams observed at each dose

level. The replicated count responses in conjunction with a nonparametric mixture of Binomial distributions, induced at each dose value by the DDP mixture model, enable flexible inference for the response distribution at any observed dose level. And, importantly, the dependence of the DDP prior across dose levels allows data-driven prediction for collections of response distributions, as well as inference for the implied dose-response relationship, through interpolation (and extrapolation) over any range of dose values of interest. We develop properties of the DDP model that are key for the application to developmental toxicity studies. We discuss various forms of inference that are available under the model, and design a Markov chain Monte Carlo (MCMC) posterior simulation method to implement such inference. Traditional parametric dose-response models are shown to be special (limiting) cases of the nonparametric DDP mixture model, which, using simulated data sets, is also compared with simpler semiparametric Bayesian methods. In particular, in the context of the simulation study, we provide comparison of the semiparametric model from Dominici and Parmigiani (2001) with the proposed DDP model.

The outline of the paper is as follows. Section 2 develops the DDP mixture model, including study of model properties and of the dose-response relationship, and methods for prior specification, MCMC posterior simulation and risk assessment inference. Section 3 introduces a simulation study to test the performance of the model and to compare with alternative semiparametric Bayesian approaches. In Section 4, we present applications to the 2,4,5-T and DEHP data. Section 5 concludes with discussion, including possible extensions of the methodology.

## 2 Methods

### 2.1 DDP Binomial mixture model

Under the Segment II toxicity study design, exposure occurs after implantation. Thus, following standard arguments from the literature (e.g., Zhu et al., 1994), we treat the number of implants,  $m$ , as a random quantity containing no information about the dose-response relationship. That is, we assume  $m \mid \boldsymbol{\kappa} \sim f(m; \boldsymbol{\kappa})$ , where  $\boldsymbol{\kappa}$  are parameters of the implant distribution, which do

not depend on  $x$ . Here, we assume a shifted Poisson distribution with support on  $m \geq 1$ , that is,  $f(m; \lambda) = e^{-\lambda} \lambda^{m-1} / (m-1)!$ , although more flexible distributions can be readily utilized. We focus on the data structure that involves the number of implants,  $m_{ij}$ , and the corresponding number of combined negative outcomes,  $y_{ij}$ , for dam  $j = 1, \dots, n_i$  at dose level  $x_i$ ,  $i = 1, \dots, N$ . (In Section 5, we discuss the model extension for the multcategory classification  $(m_{ij}, R_{ij}, y'_{ij})$  considered in Section 1.2.) With the assumption of an implant distribution that does not depend on dose level, the modeling for the number of implants and the number of negative outcomes is decomposed to  $f(m, y) = f(m)f(y | m)$ . Therefore, inference for the parameters of the implant distribution is carried out separately from inference for the parameters of the model for  $f(y | m)$ .

We propose nonparametric mixture modeling for the response distribution given the number of implants, using an extension of Dirichlet process (DP) mixing to incorporate the dependence on the dose level. We use  $\text{DP}(\alpha, G_0)$  to denote the DP prior (Ferguson, 1973) defined in terms of a centering (base) distribution  $G_0$ , and precision parameter  $\alpha > 0$ . Using its constructive definition (Sethuraman, 1994), the DP generates countable mixtures of point masses with locations drawn from the base distribution and weights defined by a stick-breaking process. Specifically, a random distribution,  $G$ , drawn from  $\text{DP}(\alpha, G_0)$  has an almost sure representation as  $G(\cdot) = \sum_{l=1}^{\infty} \omega_l \delta_{\eta_l}(\cdot)$ , where  $\delta_a$  denotes a point mass at  $a$ , the  $\eta_l$  are i.i.d. from  $G_0$ , and  $\omega_1 = \zeta_1$ ,  $\omega_l = \zeta_l \prod_{r=1}^{l-1} (1 - \zeta_r)$  for  $l \geq 2$ , with  $\zeta_l$  i.i.d. from a  $\text{Beta}(1, \alpha)$  distribution (independently of the  $\eta_l$ ).

To achieve modeling for the response distribution that allows nonparametric dependence structure across dose levels, we represent  $f(y | m)$  as a mixture of Binomial distributions with dose-dependent mixing distribution. Placing a DDP prior on the collection of mixing distributions  $\{G_x : x \in \mathcal{X}\}$  (where  $\mathcal{X} \subseteq \mathbb{R}^+$ ), which are indexed by dose level  $x$ , yields the desired nonparametric prior model for the collection of dose-dependent response distributions.

To define the DDP prior used in this paper, the (almost sure) DP representation is extended to

$$G_{\mathcal{X}}(\cdot) = \sum_{l=1}^{\infty} \omega_l \delta_{\eta_{lx}}(\cdot) \quad (1)$$

where the  $\eta_{l\mathcal{X}} = \{\eta_l(x) : x \in \mathcal{X}\}$  are i.i.d. realizations from a stochastic process  $G_{0\mathcal{X}}$  over  $\mathcal{X}$ . A key feature of the DDP prior is that for any finite collection of dose levels  $(x_1, \dots, x_k)$  it induces a multivariate DP prior for the corresponding collection of mixing distributions  $(G_{x_1}, \dots, G_{x_k})$ . Therefore, the DDP prior model involves a countable mixture of realizations from stochastic process  $G_{0\mathcal{X}}$  with weights matching those from the standard DP; this prior structure is referred to as “single- $p$ ” DDP prior (MacEachern, 2000; DeIorio et al., 2004; Gelfand et al., 2005; Kottas et al., 2008; Rodriguez and ter Horst, 2008; Kottas and Krnjajić, 2009).

Hence, we propose the following DDP prior mixture model

$$f(y | m; G_{\mathcal{X}}) = \int \text{Bin} \left( y; m, \frac{\exp(\theta)}{1 + \exp(\theta)} \right) dG_{\mathcal{X}}(\theta), \quad G_{\mathcal{X}} | \alpha, \psi \sim \text{DDP}(\alpha, G_{0\mathcal{X}}) \quad (2)$$

where  $\text{DDP}(\alpha, G_{0\mathcal{X}})$  denotes the DDP prior for  $G_{\mathcal{X}} = \sum_{l=1}^{\infty} \omega_l \delta_{\eta_{l\mathcal{X}}}$  with precision parameter  $\alpha$  and base stochastic process  $G_{0\mathcal{X}}$  that depends on parameters  $\psi$ ; the full Bayesian model is implemented with hyperpriors on  $\alpha$  and  $\psi$ . The logistic transformation for the probability of the Binomial kernel is used to facilitate the DDP prior formulation, in particular, we take  $G_{0\mathcal{X}}$  to be a Gaussian process (GP) with a linear mean function, constant variance, and isotropic power exponential correlation function. Therefore, for all  $l$ ,  $\text{E}(\eta_l(x) | \beta_0, \beta_1) = \beta_0 + \beta_1 x$ ,  $\text{Var}(\eta_l(x) | \sigma^2) = \sigma^2$ , and  $\text{Corr}(\eta_l(x), \eta_l(x') | \phi) = \exp(-\phi|x - x'|^d)$ , with  $\phi > 0$  and (fixed)  $d \in [1, 2]$  (and thus  $\psi = (\beta_0, \beta_1, \sigma^2, \phi)$ ). The linear mean function is key for flexible inference about the dose-response relationship implied by model (2) (see Section 2.3). Moreover, as discussed in Section 2.2, it enables connections with standard parametric dose-response models, which arise as limiting cases of the DDP mixture model.

Regarding nonparametric Binomial mixtures, an early reference is Berry and Christensen (1979), where the following hierarchical model was considered:  $y_i | \pi_i \stackrel{\text{ind.}}{\sim} \text{Bin}(y_i; m_i, \pi_i)$ ;  $\pi_i | G \stackrel{\text{iid}}{\sim} G$ , for  $i = 1, \dots, n$ ; and  $G \sim \text{DP}(\alpha, G_0)$ , with fixed  $\alpha$  and  $G_0$ , defined through a Beta distribution. They obtained expressions for  $\text{E}(G | \text{data})$  for  $n \leq 3$ , and approximations to  $\text{E}(G | \text{data})$  and  $\text{E}(\pi_i | \text{data})$  for large  $n$ . Under the same model, albeit with random  $\alpha$ ,

Liu (1996) developed sequential imputation techniques for empirical Bayesian inference (with  $\alpha$  estimated by its MLE) on: the number of clusters among the  $\pi_i$ ; the posterior mean and variance of each  $\pi_i$ ; and the posterior mean of  $G$ . More recently, Zhang and Liu (2012) studied an alternative nonparametric Binomial mixture model, using a Bernstein-DP prior for  $G$ . Model (2) adds on existing methods for nonparametric hierarchical modeling of Binomial data, as it involves dependent Binomial mixtures indexed by values in the (uncountable) space  $\mathcal{X}$ .

## 2.2 Model properties

Hereinafter,  $\pi(u) = \exp(u)/(1 + \exp(u))$ ,  $u \in \mathbb{R}$ , will be used to denote the logistic function.

The DDP Binomial mixture model in (2) includes both the hierarchical Binomial-logistic-normal model and the standard Binomial-logit model as special (limiting) cases. As  $\alpha \rightarrow \infty$ , each response replicate has a distinct mixing parameter. If we also assume the GP for  $G_{0\mathcal{X}}$  is a white noise process, we obtain the hierarchical Binomial-logistic-normal model, that is,  $y_{ij} \mid m_{ij}, \theta_{ij} \stackrel{ind.}{\sim} \text{Bin}(y_{ij}; m_{ij}, \pi(\theta_{ij}))$ , with  $\theta_{ij} \mid \beta_0, \beta_1, \sigma^2 \stackrel{ind.}{\sim} \text{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ . If we let  $\sigma^2 \rightarrow 0^+$ , we arrive at the standard Binomial-logit model as a further special limiting case. In the other extreme, as  $\alpha \rightarrow 0^+$ , all the response replicates are assigned to a single mixture component. This, along with the white noise process assumption, yields a Binomial-logistic normal model with a common mean for each animal within a dose level, that is,  $y_{ij} \mid m_{ij}, \theta_i \stackrel{ind.}{\sim} \text{Bin}(y_{ij}; m_{ij}, \pi(\theta_i))$ , with  $\theta_i \mid \beta_0, \beta_1, \sigma^2 \stackrel{ind.}{\sim} \text{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ . Again, with the additional restriction of  $\sigma^2 \rightarrow 0^+$ , we obtain the standard Binomial-logit model.

The DDP prior for  $G_{\mathcal{X}} = \sum_{l=1}^{\infty} \omega_l \delta_{\eta_{l\mathcal{X}}}$  allows for a flexible response distribution at each dose level through a DP mixture of Binomial distributions, induced by the mixture model in (2). Consider a realization  $\theta_{\mathcal{X}} = \{\theta(x) : x \in \mathcal{X}\}$ , which, given  $G_{\mathcal{X}}$ , arises from  $G_{\mathcal{X}}$ . Then, for any  $x, x' \in \mathcal{X}$ ,  $\text{Cov}(\theta(x), \theta(x') \mid G_{\mathcal{X}}) = \sum \omega_l \eta_l(x) \eta_l(x') - \{\sum \omega_l \eta_l(x)\} \{\sum \omega_l \eta_l(x')\}$ . Therefore, although  $G_{\mathcal{X}}$  is centered around a GP with isotropic covariance function, it generates nonstationary realizations with non-Gaussian finite dimensional distributions. Moreover, if  $G_x$  and  $G_{x'}$  denote the marginal distributions of  $\theta(x)$  and  $\theta(x')$  under  $G_{\mathcal{X}}$ , then the continuity of the  $\eta_{l\mathcal{X}}$  implies

that, as the distance between  $x$  and  $x'$  gets smaller, the difference between  $G_x$  and  $G_{x'}$  gets smaller; formally, for any  $\varepsilon > 0$ ,  $\lim_{|x-x'|\rightarrow 0} \Pr(\mathcal{L}(G_x, G_{x'}) < \varepsilon) = 1$ , where  $\mathcal{L}$  is the Lévy distance (MacEachern, 2000). Hence, the level of dependence between  $G_x$  and  $G_{x'}$ , and thus between  $f(y | m; G_x)$  and  $f(y | m; G_{x'})$ , is driven by the distance of the dose levels. The practical implication is that in prediction for the probability mass function  $f(y | m; G_x)$  and for the corresponding dose-response curve, we learn more from dose levels  $x'$  nearby  $x$  than from more distant dose levels, a desirable property for distributions that are expected to evolve relatively smoothly with the dose level.

In studying certain properties of the DDP mixture model it is convenient to apply a finite truncation approximation to  $G_{\mathcal{X}}$  in (1), and this will also provide the basis for MCMC posterior simulation (see Section 2.4.1). Specifically, we will work with  $G_{\mathcal{X}}^L = \sum_{l=1}^L p_l \delta_{Z_{l\mathcal{X}}}$ , where the  $Z_{l\mathcal{X}} = \{Z_l(x) : x \in \mathcal{X}\}$  are i.i.d. realizations, given  $\psi$ , from  $G_{0\mathcal{X}}$ , and the weights  $p_l$  arise from a truncated version of the stick-breaking construction:  $p_1 = V_1$ ,  $p_l = V_l \prod_{r=1}^{l-1} (1 - V_r)$ ,  $l = 2, \dots, L-1$ , and  $p_L = 1 - \sum_{l=1}^{L-1} p_l$ , with the  $V_l$  i.i.d., given  $\alpha$ , from  $\text{Beta}(1, \alpha)$ . The truncation level can be chosen using standard distributional properties for the weights arising from the stick-breaking structure in (1). For instance,  $E(\sum_{l=1}^L \omega_l | \alpha) = 1 - \{\alpha/(\alpha+1)\}^L$ , which can be averaged over the prior for  $\alpha$  to estimate  $E(\sum_{l=1}^L \omega_l)$ . Given a tolerance level for the approximation, this expression is solved numerically to obtain the corresponding value  $L$ .

Next, we discuss a useful connection of the DDP mixture model in (2), which is built from the Binomial kernel for the number of combined negative outcomes within a dam, with a DDP mixture model based on a product of Bernoulli kernel for the set of binary responses for all implants corresponding to that dam. The mixture model using the underlying vector of binary responses,  $\mathbf{y}^* = (y_1^*, \dots, y_m^*)$ , for a generic dam with  $m$  implants at dose level  $x$  is given by

$$f^*(\mathbf{y}^* | m; G_{\mathcal{X}}) = \int \prod_{k=1}^m \text{Bern}(y_k^*; \pi(\theta)) dG_{\mathcal{X}}(\theta), \quad (3)$$

where the same DDP prior as before would be assigned to  $G_{\mathcal{X}}$ . Note that the model formulation

involves a common mixing parameter for all binary outcomes associated with the same dam. Mixture models (2) and (3) are equivalent in the sense that the moment generating function for the number of negative outcomes under (2),  $E(e^{ty} | m; G_{\mathcal{X}})$ , can be straightforwardly shown to be equal to the moment generating function for the sum of binary responses under (3),  $E(e^{t \sum_{k=1}^m y_k^*} | m; G_{\mathcal{X}})$ . This result is used to study the dose-response curve implied by the DDP Binomial mixture model (see Section 2.3).

We can also determine the correlation between two binary responses within the same dam at a generic dose level  $x$ , i.e.,  $\text{Corr}(y_k^*, y_{k'}^*; G_x^L)$ ; we will refer to this as the intracluster correlation (where the dam serves as the *cluster*). Suppressing the implicit conditioning on  $m = 1$  or  $m = 2$ , the expectations needed to obtain the intracluster correlation are  $E(y_k^*; G_x^L) = E(y_{k'}^*; G_x^L) = \sum_{l=1}^L p_l \pi(Z_l(x))$ , and  $E(y_k^* y_{k'}^*; G_x^L) = \sum_{l=1}^L p_l (\pi(Z_l(x)))^2$ . Moreover,  $\text{Var}(y_k^*; G_x^L) = \text{Var}(y_{k'}^*; G_x^L) = \{\sum_{l=1}^L p_l \pi(Z_l(x))\} - \{\sum_{l=1}^L p_l \pi(Z_l(x))\}^2$ , and thus

$$\text{Corr}(y_k^*, y_{k'}^*; G_x^L) = \frac{\{\sum_{l=1}^L p_l (\pi(Z_l(x)))^2\} - \{\sum_{l=1}^L p_l \pi(Z_l(x))\}^2}{\{\text{Var}(y_k^*; G_x^L) \text{Var}(y_{k'}^*; G_x^L)\}^{1/2}}. \quad (4)$$

Developmental toxicity studies typically give rise to overdispersed binary responses, that is,  $\text{Corr}(y_k^*, y_{k'}^*) > 0$ . Capitalizing on results for mixtures of exponential families, it can be shown (see the Appendix) that the DDP mixture model supports positive intracluster correlations.

### 2.3 Dose-response relationship

Using the mixture model formulation in (3) for the underlying binary outcomes, we define the dose-response curve through the probability of a negative outcome for a generic implant expressed as a function of dose level. Therefore, under the DDP truncation approximation discussed in Section 2.2, the dose-response curve is given by

$$\Pr(y^* = 1; G_{\mathcal{X}}^L) = \int \pi(\theta) dG_{\mathcal{X}}^L(\theta) = \sum_{l=1}^L p_l \pi(Z_{l\mathcal{X}}). \quad (5)$$

Note that, although this is a conditional probability (given  $m = 1$ ), we again suppress this implicit conditioning in the notation.

Smoothness properties of prior realizations for the dose-response function emerge directly from properties of prior realizations  $Z_{l\mathcal{X}}$  under the centering GP  $G_{0\mathcal{X}}$ . In particular, for choices of  $d \in [1, 2)$  ( $d = 2$ ) for the GP correlation function, the continuity (differentiability) of the  $Z_{l\mathcal{X}}$  yields continuous (differentiable) dose-response curves under the DDP Binomial mixture model.

A key aspect of the model is that it does not force a monotonicity restriction to the dose-response function, which is an assumption for standard parametric dose-response models. However, the prior expectation  $E(\Pr(y^* = 1; G_{\mathcal{X}}^L))$  is non-decreasing with  $x$  provided  $\beta_1 > 0$  (see the Appendix), and this is crucial for practicable posterior inference. In particular, if the model is applied using a constant mean function for the DDP prior centering GP (i.e., setting  $\beta_1 = 0$ ), there is little hope to obtain meaningful interpolation and extrapolation results for the dose-response curve. Even though we insist on the non-decreasing trend in its prior expectation, prior (and thus posterior) realizations for the dose-response curve are not structurally restricted to be non-decreasing. An illustration of this model feature is provided in Section 4.2, where the shape of the estimated dose-response curve for the DEHP data is indicative of a possible hormetic dose-response relationship, as discussed in Section 1.2. The fact that the DDP mixture model allows non-monotonic dose-response relationships to be uncovered is an asset of the proposed modeling approach, and, arguably, a practical advance relative to existing methods.

## 2.4 Implementation details for posterior inference

### 2.4.1 Hierarchical model and MCMC posterior simulation

To motivate the hierarchical model formulation for the data, we note that for the DEHP data, discussed in Section 1.2, the dams are labeled and recorded in ascending numerical order across dose levels; that is, the smallest ID number corresponds to data from the first dam at the first dose level, the first dam at the second dose level has the next ID number, and so on. (This is also the case for other data sets available from the database of the National Toxicology

Program.) Therefore, the animals can be linked as a response vector across the dose levels with the conditional independence assumption built for the replicated response vectors. Hence, the data structure and corresponding hierarchical model is along the lines of the spatial DP (Gelfand et al., 2005) rather than, for instance, the ANOVA DDP (DeIorio et al., 2004). Nevertheless, for the data sets we considered results did not differ significantly under the alternative hierarchical model for the data based on exchangeability both across and within dose levels.

More specifically, let  $\mathbf{y}_j = (y_{1j}, \dots, y_{Nj})$  be the  $j$ th response replicate with number of implants vector  $\mathbf{m}_j = (m_{1j}, \dots, m_{Nj})$ , for  $j = 1, \dots, n$  (where  $n = \max_i n_i$ ), and  $\boldsymbol{\theta}_j \equiv \boldsymbol{\theta}_j(\mathbf{x}) = (\theta_j(x_1), \dots, \theta_j(x_N))$  be the latent mixing vector for  $\mathbf{y}_j$ , where  $\mathbf{x} = (x_1, \dots, x_N)$ . We introduce missing value indicators,  $s_{ij}$ , such that  $s_{ij} = 1$  if the  $j$ th replicate at dose level  $i$  is present and  $s_{ij} = 0$  otherwise. (Note that the  $s_{ij}$  are fully specified for any particular data set.) Then, the first stage of the hierarchical model for the data on combined negative outcomes is written as  $\{y_{ij}\} \mid \{m_{ij}\}, \{\boldsymbol{\theta}_j\} \sim \prod_{j=1}^n \prod_{i=1}^N \{\text{Bin}(y_{ij}; m_{ij}, \pi(\theta_j(x_i)))\}^{s_{ij}}$ . Here, the  $\boldsymbol{\theta}_j$ , given  $G_{\mathbf{x}}$ , are i.i.d.  $G_{\mathbf{x}}$ , and  $G_{\mathbf{x}}$  has a DP( $\alpha, G_{0\mathbf{x}}$ ) prior implied by the DDP prior for  $G_{\mathcal{X}}$ . In particular,  $G_{0\mathbf{x}} = N_N((\beta_0 + \beta_1 x_1, \dots, \beta_0 + \beta_1 x_N)^T, \boldsymbol{\Sigma})$ , where  $\boldsymbol{\Sigma}$  is induced by the GP covariance function, that is,  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{H}(\phi)$  with  $\mathbf{H}_{ij}(\phi) = \exp(-\phi|x_i - x_j|^d)$ .

Hence, the hierarchical model for the data is a DP mixture model induced by the DDP mixture prior. For MCMC posterior simulation, we use blocked Gibbs sampling (e.g., Ishwaran and James, 2001) based on truncation of  $G_{\mathbf{x}}$ , which is induced by the finite truncation approximation to  $G_{\mathcal{X}}$  discussed in Section 2.2. Although other MCMC methods could be considered, we are drawn to the ready implementation of the blocked Gibbs sampler, and the ease with which it can handle unbalanced response replicates.

Truncating  $G_{\mathbf{x}}$  at a sufficiently large level  $L$ , the model includes  $G_{\mathbf{x}}^L \equiv (\mathbf{p}, \mathbf{Z}) = \sum_{l=1}^L p_l \delta_{\mathbf{Z}_l(\mathbf{x})}$ , where the weights  $\mathbf{p} = (p_1, \dots, p_L)$  are defined in Section 2.2, and  $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_L)$ , with  $\mathbf{Z}_l \equiv \mathbf{Z}_l(\mathbf{x}) = (Z_l(x_1), \dots, Z_l(x_N))$ . Hence, under the truncated version of mixing distribution  $G_{\mathbf{x}}$ ,  $\boldsymbol{\theta}_j = \mathbf{Z}_l$  with probability  $p_l$ . Introducing configuration variables  $\mathbf{w} = (w_1, \dots, w_n)$ , where each  $w_j$  takes a value in  $\{1, \dots, L\}$ , we have  $w_j = l$  if and only if  $\boldsymbol{\theta}_j = \mathbf{Z}_l$ , for  $l = 1, \dots, L$  and

$j = 1, \dots, n$ . Then, we can write the hierarchical model for the  $y_{ij}$  as

$$\begin{aligned} \{y_{ij}\} \mid \{m_{ij}\}, \mathbf{w}, \mathbf{Z} &\sim \prod_{j=1}^n \prod_{i=1}^N \left\{ \text{Bin}(y_{ij}; m_{ij}, \pi(Z_{w_j}(x_i))) \right\}^{s_{ij}} \\ w_j \mid \mathbf{p} &\stackrel{i.i.d.}{\sim} \sum_{l=1}^L p_l \delta_l(w_j), \quad j = 1, \dots, n \end{aligned} \quad (6)$$

where, conditionally on  $\boldsymbol{\psi}$ , the  $\mathbf{Z}_l(\mathbf{x})$ ,  $l = 1, \dots, L$ , are i.i.d. from  $G_{0\mathbf{x}}$ , and the prior density for  $\mathbf{p}$  is given by  $\alpha^{L-1} p_L^{\alpha-1} (1-p_1)^{-1} (1-(p_1+p_2))^{-1} \times \dots \times (1-\sum_{l=1}^{L-2} p_l)^{-1}$ , which is a special case of the generalized Dirichlet distribution. The model is completed with (independent) priors for the DDP hyperparameters, in particular, we place a  $\text{gamma}(a_\alpha, b_\alpha)$  prior on  $\alpha$ , a  $\text{N}(m_0, s_0^2)$  prior on  $\beta_0$ , an inverse gamma prior on  $\sigma^2$  with shape parameter  $a_\sigma > 1$  and mean  $b_\sigma / (a_\sigma - 1)$ , and a uniform prior on  $\phi$  over  $(0, b_\phi)$ . Moreover, an exponential prior is taken for  $\beta_1$  to incorporate the non-decreasing trend in the prior expectation for the dose-response curve.

Denote the  $n^*$  distinct values of vector  $\mathbf{w}$  by  $w_1^*, \dots, w_{n^*}^*$ , and let  $M_k^* = |\{j : w_j = w_k^*\}|$ , for  $k = 1, \dots, n^*$ , and  $M_l = |\{w_j : w_j = l\}|$ , for  $l = 1, \dots, L$ . Then, sampling from the posterior distribution  $p(\mathbf{p}, \mathbf{Z}, \mathbf{w}, \alpha, \boldsymbol{\psi} \mid \text{data})$  corresponding to model (6) is based on simulation from the following posterior full conditional distributions.

**Z update:** The full conditional for  $\mathbf{Z}_l$  depends on whether  $l$  corresponds to one of the distinct components. If  $l \notin \{w_k^* : k = 1, \dots, n^*\}$ , then  $\mathbf{Z}_l$  is drawn from  $\text{N}_N(\beta_0 \mathbf{j}_N + \beta_1 \mathbf{x}, \boldsymbol{\Sigma})$ , where  $\mathbf{j}_N$  denotes a vector of dimension  $N$  with all elements equal to 1. For  $l \in \{w_k^* : k = 1, \dots, n^*\}$ ,

$$\mathbf{Z}_{w_k^*} \mid \mathbf{w}, \boldsymbol{\psi}, \text{data} \propto \text{N}_N(\mathbf{Z}_{w_k^*}; \beta_0 \mathbf{j}_N + \beta_1 \mathbf{x}, \boldsymbol{\Sigma}) \prod_{\{j:w_j=w_k^*\}} \prod_{i=1}^N \left\{ \text{Bin}(y_{ij}; m_{ij}, \pi(Z_{w_k^*}(x_i))) \right\}^{s_{ij}}$$

which is sampled with a Metropolis-Hastings step. We use an  $N$ -variate Gaussian proposal distribution, centered at the previous iteration. The covariance matrix of the proposal distribution is of the same form as the GP prior,  $\mathbf{D}_{ij} = a \exp(-b|x_i - x_j|)$ , where  $a$  and  $b$  are tuning parameters. The acceptance rates for the data sets of Sections 3 and 4 were between 0.15 and 0.20.

**w update:** The conditional posterior for each  $w_j$ ,  $j = 1, \dots, n$ , is a discrete distribution,

$\sum_{l=1}^L \tilde{p}_{lj} \delta_l(w_j)$ , where  $\tilde{p}_{lj} \propto p_l \prod_{i=1}^N \{\text{Bin}(y_{ij}; m_{ij}, \pi(Z_l(x_i)))\}^{s_{ij}}$ , for  $l = 1, \dots, L$ .

**p update:** The conditional posterior of  $\mathbf{p}$  is given by a generalized Dirichlet distribution with parameters  $(M_1+1, \dots, M_{L-1}+1)$  and  $(\alpha + \sum_{k=2}^L M_k, \dots, \alpha + M_L)$ . Hence, vector  $\mathbf{p}$  can be sampled by generating latent  $V_l^*$ ,  $l = 1, \dots, L-1$ , independently from  $\text{Beta}(M_l+1, \alpha + \sum_{k=l+1}^L M_k)$ , and setting  $p_1 = V_1^*$ ,  $p_l = V_l^* \prod_{r=1}^{l-1} (1 - V_r^*)$ ,  $l = 2, \dots, L-1$ , and  $p_L = 1 - \sum_{l=1}^{L-1} p_l = \prod_{r=1}^{L-1} (1 - V_r^*)$ .

**Hyperparameter updates:** The full conditional for  $\alpha$  is a gamma distribution with shape parameter  $L + a_\alpha - 1$  and rate parameter  $b_\alpha - \log p_L = b_\alpha - \sum_{l=1}^{L-1} \log(1 - V_l^*)$ . The posterior full conditional of  $\beta_0$  is normal with mean  $(m_0 s_0^{-2} + \mathbf{j}'_N \Sigma^{-1} \sum_{k=1}^{n^*} (\mathbf{Z}_{w_k^*} - \beta_1 \mathbf{x})) / (s_0^{-2} + n^* \mathbf{j}'_N \Sigma^{-1} \mathbf{j}_N)$  and variance  $(s_0^{-2} + n^* \mathbf{j}'_N \Sigma^{-1} \mathbf{j}_N)^{-1}$ . Moreover,  $\sigma^2$  has an inverse gamma full conditional with shape parameter  $a_\sigma + 0.5 n^* N$  and rate parameter  $b_\sigma + 0.5 \sum_{k=1}^{n^*} (\mathbf{Z}_{w_k^*} - \beta_0 \mathbf{j}_N - \beta_1 \mathbf{x})' \mathbf{H}^{-1}(\phi) (\mathbf{Z}_{w_k^*} - \beta_0 \mathbf{j}_N - \beta_1 \mathbf{x})$ . Parameter  $\beta_1$  is updated with a random-walk Metropolis-Hastings step based on a normal proposal distribution on the logarithmic scale. And, parameter  $\phi$  is sampled by discretizing its bounded support, induced by the  $\text{Unif}(0, b_\phi)$  prior.

Finally, note that inference for the implant distribution, using a gamma prior for  $\lambda$ , is implemented independent of the DDP mixture, and is not discussed further except with regard to posterior predictive calculations in the next section.

## 2.4.2 Inference for the dose-response relationship and risk assessment

Consider new responses  $(m_0, y_0)$  corresponding to a generic dose level  $x_0$ . Under the assumed formulation for the joint distribution,  $f(m, y; \lambda, G_{\mathcal{X}}) = f(m; \lambda) f(y | m; G_{\mathcal{X}})$ , the posterior predictive distribution for  $(m_0, y_0)$  can be separated into the conditional predictive for  $y_0$  and the marginal predictive for  $m_0$ . That is,  $p(m_0, y_0 | x_0, \text{data}) = \int f(m_0; \lambda) p(\lambda | \text{data}) d\lambda \times p(y_0 | m_0, x_0, \text{data})$ . The expression for  $p(y_0 | m_0, x_0, \text{data})$  depends on whether  $x_0$  is among the observed doses or a new dose level, and can be obtained as a special case of the general form below. Therefore, to obtain inference for the joint posterior predictive distribution, at each iteration of the MCMC algorithm, we draw  $m_0$  from a shifted Poisson with mean  $\lambda$ , then given  $m_0$  and the current values of the DDP parameters, obtain a predictive draw for  $y_0$ .

Moreover, each posterior sample for  $(\mathbf{p}, \mathbf{Z})$  provides a posterior realization for  $G_{\mathbf{x}}^L$  directly through its definition,  $\sum_{l=1}^L p_l \delta_{\mathbf{Z}_l(\mathbf{x})}$ . Next, given the predictive draw for the number of implants  $m_0$ , for any vector  $\mathbf{y}_0 = (y_{10}, \dots, y_{N0})$ ,  $f(\mathbf{y}_0 | m_0; G_{\mathbf{x}}^L) = \sum_{l=1}^L p_l \prod_{i=1}^N \text{Bin}(y_{i0}; m_0, \pi(Z_l(x_i)))$  is a posterior realization from the conditional response distribution at the observed doses.

To extend the inference beyond the  $N$  observed dose levels, we predict across  $M$  new doses,  $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_M)$ , which may include values outside the range of the observed doses. To predict a new vector of responses at all dose levels,  $(\mathbf{y}_0, \tilde{\mathbf{y}}_0) = (y_{10}, \dots, y_{N0}, \tilde{y}_{10}, \dots, \tilde{y}_{M0})$ , given the corresponding number of implants  $m_0$ , the mixing parameter vector is extended to  $(\mathbf{Z}_l(\mathbf{x}), \tilde{\mathbf{Z}}_l(\tilde{\mathbf{x}}))$ , for  $l = 1, \dots, L$ , where given model parameters and the  $\mathbf{Z}_l(\mathbf{x})$ , the  $\tilde{\mathbf{Z}}_l(\tilde{\mathbf{x}})$  are obtained through standard conditioning under multivariate normal distributions. Denoting  $\tilde{\mathbf{Z}} = \{\tilde{\mathbf{Z}}_l(\tilde{\mathbf{x}}) : l = 1, \dots, L\}$ , the conditional predictive distribution for  $(\mathbf{y}_0, \tilde{\mathbf{y}}_0)$  given  $m_0$  is given by

$$p((\mathbf{y}_0, \tilde{\mathbf{y}}_0) | m_0, \tilde{\mathbf{x}}, \text{data}) = \int \int \sum_{l=1}^L p_l \left\{ \prod_{i=1}^N \text{Bin}(y_{i0}; m_0, \pi(Z_l(x_i))) \prod_{j=1}^M \text{Bin}(\tilde{y}_{j0}; m_0, \pi(\tilde{Z}_l(\tilde{x}_j))) \right\} \times \left( \prod_{l=1}^L N_M(\tilde{\mathbf{Z}}_l(\tilde{\mathbf{x}}); \tilde{\boldsymbol{\mu}}_l, \tilde{\boldsymbol{\Sigma}}) \right) d\tilde{\mathbf{Z}} d\mathbf{p}(\mathbf{p}, \mathbf{Z}, \alpha, \boldsymbol{\psi} | \text{data}).$$

Here,  $\tilde{\boldsymbol{\mu}}_l = (\beta_0 \mathbf{j}_M + \beta_1 \tilde{\mathbf{x}}) + \mathbf{H}^{MN}(\phi) \mathbf{H}^{-1}(\phi) (\mathbf{Z}_l(\mathbf{x}) - (\beta_0 \mathbf{j}_N + \beta_1 \mathbf{x}))$ , and  $\mathbf{H}^{MN}(\phi)$  is the  $M \times N$  matrix with  $\mathbf{H}_{ij}^{MN}(\phi) = \exp(-\phi |\tilde{x}_i - x_j|^d)$ . Moreover,  $\tilde{\boldsymbol{\Sigma}} = \sigma^2 \{ \mathbf{H}^{MM}(\phi) - \mathbf{H}^{MN}(\phi) \mathbf{H}^{-1}(\phi) (\mathbf{H}^{MN}(\phi))^T \}$ , where  $\mathbf{H}^{MM}(\phi)$  is the  $M \times M$  matrix with  $\mathbf{H}_{ij}^{MM}(\phi) = \exp(-\phi |\tilde{x}_i - \tilde{x}_j|^d)$ .

Using the posterior samples for the parameters, we can evaluate expression (4) to develop inference for the intracluster correlation as a function of dose. Standard hierarchical extensions of the Binomial model are limited with regard to such inference, e.g., the Beta-binomial model involves the same (positive) correlation for all dose levels. The data of Section 4.1 illustrate the capacity of the DDP mixture model to uncover dose-dependent intracluster correlation patterns.

Key to quantitative risk assessment is inference for the dose-response relationship. Using the posterior samples for  $(\mathbf{p}, \mathbf{Z}, \tilde{\mathbf{Z}})$ , we obtain the posterior distribution of  $\Pr(y^* = 1; G_{x_0}^L) = \sum_{l=1}^L p_l \pi(Z_l(x_0))$  arising from (5) for each  $x_0 \in (\mathbf{x}, \tilde{\mathbf{x}})$ . These distributions can be summarized

with posterior means and two percentiles to provide point and interval estimates for the dose-response curve (as in Figures 8 and 9). It can be readily shown that the posterior expectation of  $\Pr(y^* = 1; G_{x_0}^L)$  is equal to the expectation of  $y_0/m_0$  from the joint posterior predictive distribution  $p(m_0, y_0 \mid x_0, \text{data})$ , that is,  $E(y_0/m_0 \mid \text{data}) = E(\Pr(y^* = 1; G_{x_0}^L) \mid \text{data})$ . As shown in Figure 5, the posterior predictive samples for  $y_0/m_0$  obtained across a range of dose levels also provide useful inference for the dose-response relationship.

The expression of the posterior mean for the dose-response relationship can be elaborated to elucidate the clustering induced by the DDP mixture structure. Consider the binary response  $y^*$  for a generic implant at observed dose level  $x_0$  (for a new dose  $\tilde{x}_0$ , we need to also average in the following expressions over the normal distributions for the  $\tilde{Z}_l(\tilde{x}_0)$ ). Then,  $\Pr(y^* = 1 \mid \text{data}) \equiv E(\Pr(y^* = 1; G_{x_0}^L) \mid \text{data}) = \int \left\{ \sum_{w_0} \pi(Z_{w_0}(x_0)) \sum_{l=1}^L p_l \delta_l(w_0) \right\} dp(\mathbf{p}, \mathbf{Z}, \mathbf{w}, \alpha, \boldsymbol{\psi} \mid \text{data})$ . Recall from Section 2.4.1 that the vector of mixture weights  $\mathbf{p}$  can be effectively replaced in the joint posterior by the vector of latent parameters  $\mathbf{V}^* = \{V_l^* : l = 1, \dots, L-1\}$ , the posterior full conditional distribution of which is given by  $p(\mathbf{V}^* \mid \mathbf{Z}, \mathbf{w}, \alpha, \boldsymbol{\psi}, \text{data}) \equiv p(\mathbf{V}^* \mid \mathbf{w}, \alpha, \text{data}) = \prod_{l=1}^{L-1} \text{Beta}(V_l^*; M_l + 1, \alpha + \sum_{k=l+1}^L M_k)$ . Hence, we rewrite the  $p_l$  in terms of the  $V_l^*$ , that is,  $p_1 = V_1^*$ ,  $p_l = V_l^* \prod_{r=1}^{l-1} (1 - V_r^*)$ ,  $l = 2, \dots, L-1$ , and  $p_L = \prod_{r=1}^{L-1} (1 - V_r^*)$ . Then, by factorizing the joint posterior into  $p(\mathbf{V}^* \mid \mathbf{w}, \alpha, \text{data})p(\mathbf{Z}, \mathbf{w}, \alpha, \boldsymbol{\psi} \mid \text{data})$  and marginalizing over the  $V_l^*$ ,

$$E(\Pr(y^* = 1; G_{x_0}^L) \mid \text{data}) = \int \left\{ \sum_{l=1}^L q_l \pi(Z_l(x_0)) \right\} dp(\mathbf{Z}, \mathbf{w}, \alpha, \boldsymbol{\psi} \mid \text{data}).$$

Here,  $q_1 = (1 + M_1)/(1 + \alpha + n)$ ,  $q_l = \{(1 + M_l)/(1 + \alpha + S_l)\} \prod_{r=1}^{l-1} \{(\alpha + S_{r+1})/(1 + \alpha + S_r)\}$ , for  $l = 2, \dots, L-1$ , and  $q_L = 1 - \sum_{l=1}^{L-1} q_l = \prod_{r=1}^{L-1} \{(\alpha + S_{r+1})/(1 + \alpha + S_r)\}$ , where  $S_\ell = \sum_{k=\ell}^L M_k$ , for  $\ell = 1, \dots, L$  (with  $S_1 = n$ ). The weights  $q_l \equiv q_l(\alpha, \mathbf{w})$  make more explicit the partitioning structure induced by the mixture model. In particular, note that for moderately large sample sizes,  $(\alpha + S_r)/(1 + \alpha + S_r) \approx 1$ , for  $r = 2, \dots, L-1$ , and we can thus write  $q_l \approx (1 + M_l)/(1 + \alpha + n)$ , for  $l = 2, \dots, L-1$ , with  $q_1 = (1 + M_1)/(1 + \alpha + n)$ . Hence, given a particular MCMC posterior realization with the implied partition from  $\mathbf{w}$ , the active mixture components (i.e., components

with  $M_l > 0$ ) are the main contributors to the mixture representation for the dose-response curve,  $\sum_{l=1}^L q_l \pi(Z_l(x_0))$ , with corresponding weights  $q_l$  approximately proportional to the component size. The weights for the empty components (where  $M_l = 0$ ) will be typically small, but they yield positive probability for “new” components to be explored in the posterior predictive sampling. The posterior mean estimate for the dose-response curve arises by averaging the mixture form discussed above over all MCMC posterior samples.

Finally, risk assessment can also be based on estimation of the dose level  $x_q$  that corresponds to a specified probability,  $q$ , of a negative outcome, that is,  $q = \Pr(y^* = 1; G_{x_q}^L)$ . For any set of probabilities  $q$ , each posterior realization for  $\Pr(y^* = 1; G_{\mathcal{X}}^L)$  can be (numerically) inverted to obtain the posterior distribution for the corresponding calibrated dose levels  $x_q$ .

### 2.4.3 Prior specification

Regarding the base GP  $G_{0\mathcal{X}}$ , we have experimented with exponential and Gaussian correlation functions ( $d = 1$  and  $d = 2$ , respectively). For all data examples of Sections 3 and 4, inferences were largely unaffected by the particular choice. To specify  $b_\phi$ , we consider the limiting case of the DDP mixture model with  $\alpha \rightarrow 0^+$ , which corresponds to a Binomial response distribution with a GP prior for the dose-response function on the logistic scale. Then, under the exponential correlation function,  $3/\phi$  is the *range of dependence*, i.e., the distance between dose levels that yields correlation 0.05. The range is usually assumed to be a fraction of the maximum interpoint distance over the index space. Let  $D_{\max}$  be the maximum distance between observed doses. Since  $3/b_\phi < 3/\phi$ , we specify  $b_\phi$  such that  $3/b_\phi = rD_{\max}$  for a small  $r$ . For all data analyses considered, the posterior distribution for  $\phi$  was concentrated on values substantially smaller than  $b_\phi$ .

We set the prior mean for  $\beta_0$  to 0, and the shape parameter of the inverse gamma prior for  $\sigma^2$  to 2 (implying infinite prior variance). The prior variance for  $\beta_0$  and the prior means for  $\beta_1$  and  $\sigma^2$  are chosen by studying the induced prior distribution for the dose-response curve for which prior realizations can be readily sampled using (5). Specifically, under the prior choice discussed below, the prior mean for  $\Pr(y^* = 1; G_{\mathcal{X}}^L)$  begins around 0.5 with a slight increasing

trend, and the corresponding 95% interval bands are essentially spanning the  $(0, 1)$  interval.

The DDP prior precision parameter,  $\alpha$ , controls the number,  $n^*$ , of distinct mixture components (e.g., Antoniak, 1974; Escobar and West, 1995). In particular, for moderate to large sample sizes, a useful approximation to the prior expectation  $E(n^* | \alpha)$  is given by  $\alpha \log\{(\alpha+n)/\alpha\}$ . This expression can be averaged over the  $\text{gamma}(a_\alpha, b_\alpha)$  prior for  $\alpha$  to obtain  $E(n^*)$ , thus selecting  $a_\alpha$  and  $b_\alpha$  to agree with a prior guess at the expected number of distinct mixture components.

Note that this prior specification approach is fairly automatic as it only requires a range of dose values along with a reasonable prior for  $\alpha$ . In particular, since the range is comparable for all data examples in Sections 3 and 4, we used the same prior setting for all analyses: a normal prior for  $\beta_0$  with mean 0 and variance 20; an exponential prior for  $\beta_1$  with mean 0.1; an inverse gamma prior for  $\sigma^2$  with shape parameter 2 and mean 25; a uniform prior for  $\phi$  over  $(0, 10)$ ; and a  $\text{gamma}(2, 1)$  prior for  $\alpha$ . Prior sensitivity analysis revealed robust posterior inference under less and more dispersed priors.

Finally, regarding the truncation level  $L$  for the DDP prior approximation, for the data analyses in Sections 3 and 4,  $L = 50$  is used; based on the result discussed in Section 2.2,  $E(\sum_{l=1}^{50} \omega_l) \approx 0.9999593$  under the  $\text{gamma}(2, 1)$  prior for  $\alpha$ .

### 3 Simulation study

Here, we consider two synthetic data sets to check the performance of our model and to compare with simpler semiparametric and nonparametric Bayesian models.

#### 3.1 Simulated data

We work with simulated data sets generated under two distinct settings. The first (Figure 2, left panel) is based on a Binomial response distribution with a non-standard non-linear function for the dose-response curve. In particular, we define the probability of a negative outcome at dose  $x$  by  $\pi(h(x))$ , where  $h(x) = -2 + 0.04x - 0.25 \sin(2.7x) - 1.1/(1 + x^2)$ . The second simulation

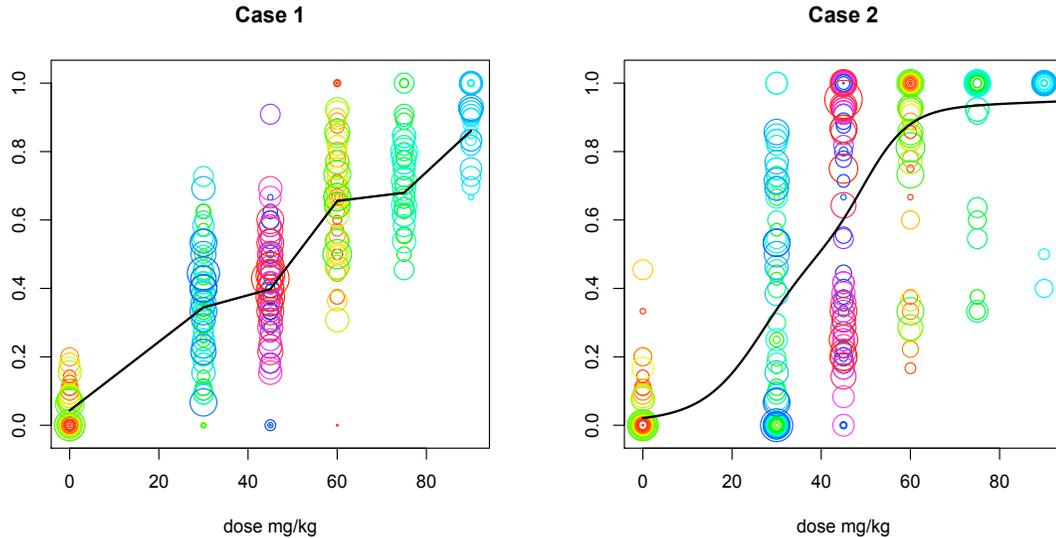


Figure 2: Simulation study. Each circle corresponds to a particular dam, the size of the circle is proportional to the number of implants, and the coordinates of the circle are the dose level and the proportion of negative outcomes. The left panel corresponds to the first simulation setting (Binomial with non-linear dose-response curve) and the right panel to the second setting (three-component mixture of Binomial-logit distributions). In each panel, the solid line denotes the true dose-response curve. The online version of this figure is in color.

example (Figure 2, right panel) is built from a mixture of three Binomial-logit distributions,  $\sum_{i=1}^3 p_i \text{Bin}(y; m, \pi(q_i(x)))$ , where  $q_1(x) = -2 + 0.02x$ ,  $q_2(x) = -10 + 0.20x$ ,  $q_3(x) = -4 + 0.15x$ , and  $(p_1, p_2, p_3) = (0.1, 0.4, 0.5)$ . For both simulations, we use the values for the dose levels, number of dams, and number of implants from the 2,4,5-T data (see Section 1.2).

### 3.2 Comparison models

For comparison, we analyze the two data sets with the semiparametric product of mixtures of Dirichlet process (PMDP) model from Dominici and Parmigiani (2001). The PMDP approach involves a different model structure than the DDP Binomial mixture. We also consider comparison with two models that can be viewed as special cases of the model developed in Section 2, a GP Binomial regression model, and a more structured DDP mixture model which ensures monotonicity for the dose-response curve.

**PMDP model:** We implement the PMDP model as in Dominici and Parmigiani (2001) with dose-specific precision parameters and a Binomial-logit centering distribution for the number of negative outcomes given a fixed number of implants. Under the PMDP model,

$$y_{ij} | F_{ij} \stackrel{ind.}{\sim} F_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, N \quad (7)$$

$$F_{ij} | \{A_i\}, (\eta_0, \eta_1) \stackrel{ind.}{\sim} \text{DP}(A_i, \text{Bin}(m_{ij}, \pi(\eta_0 + \eta_1 x_i))), \quad j = 1, \dots, n_i, \quad i = 1, \dots, N.$$

By integrating out the infinite dimensional parameters,  $F_{ij}$ , MCMC posterior sampling involves an  $N + 2$  dimensional Metropolis-Hastings step for the regression coefficients and the  $N$  dose-specific precision parameters. Conditional on the  $A_i$  and  $(\eta_0, \eta_1)$ , the posterior distribution of  $F_{ij}$  is a DP with updated parameters, and inference can be obtained using the DP definition. Following Dominici and Parmigiani (2001), we use independent normal priors for  $\eta_0$  and  $\eta_1$  with mean 0 and variance 3, and independent priors for the  $A_i$  arising from uniform distributions for the  $A_i/(10 + A_i)$ . For both simulated data sets, there is significant learning for the  $A_i$  and, especially, for  $\eta_0$  and  $\eta_1$ , under this prior choice.

The PMDP model is restrictive for inference outside the observed dose levels as the distributions are dependent in a weak fashion being linked only through the common regression coefficients. In particular, the probability of a negative outcome at a new dose is problematic to define under the version of the PMDP model in (7). Taking the precision parameter to be a function of dose, this probability can be shown to follow a  $\text{Beta}(A(x)\pi(\eta_0 + \eta_1 x), A(x)(1 - \pi(\eta_0 + \eta_1 x)))$  distribution (obtaining a result for the PMDP prior analogous to the one on the connection between DDP models (2) and (3)). Evidently, effective interpolation (or extrapolation) at new dose levels requires an appropriate dose-dependent prior model for the DP precision parameter. In general, such a specification does not seem straightforward, for instance, simple choices such as  $\log(A(x)) = \gamma_0 + \gamma_1 x$  (Carota and Parmigiani, 2002) may not be sufficiently flexible to capture the degree to which the data deviate from the centering Binomial distribution.

**GP Binomial regression model:** This model retains the restrictive Binomial response dis-

tribution, but is more flexible than the Binomial-logit model in inference for the dose-response curve. The GP model is a limiting case of the DDP mixture model (as  $\alpha \rightarrow 0^+$ ), in particular, it is based on a GP prior for the dose-response curve on the logistic scale,

$$y_{ij} \mid m_{ij}, \{\theta(x_i)\} \stackrel{ind.}{\sim} \text{Bin}(y_{ij}; m_{ij}, \pi(\theta(x_i))), \quad j = 1, \dots, n_i, \quad i = 1, \dots, N$$

where  $(\theta(x_1), \dots, \theta(x_N))$  has a  $N_N(\xi_0 \mathbf{j}_N + \xi_1 \mathbf{x}, \mathbf{\Lambda})$  prior given hyperparameters  $(\xi_0, \xi_1, \tau^2, \rho)$ . Here,  $\mathbf{\Lambda} = \tau^2 \mathbf{H}(\rho)$ , and  $\mathbf{H}_{ij}(\rho) = \exp(-\rho |x_i - x_j|^d)$ , with fixed  $d \in [1, 2]$ . We use the exponential correlation function ( $d = 1$ ) for both GP and DDP mixture models. (Inference results are similar under the Gaussian correlation function, although computing under the GP model is less stable for  $d = 2$ .) The GP hyperparameters  $(\xi_0, \xi_1, \tau^2, \rho)$  are assigned the same priors with the DDP hyperparameters  $(\beta_0, \beta_1, \sigma^2, \phi)$  given in Section 2.4.3. Note that the non-decreasing trend in prior expectation can also be incorporated to the GP model provided  $\xi_1 > 0$ .

**Linear DDP mixture model:** A distinguishing feature of the DDP mixture model of Section 2 is that it supports non-monotonic dose-response relationships. We argue that this is practically relevant in the analysis of developmental toxicology data.

However, if one wishes to enforce monotonicity for the dose-response curve (with prior probability 1 rather than only in prior expectation), this can be accomplished within the DDP mixture framework using a simplified version of the DDP prior. Specifically, setting  $\eta_l(x) = \gamma_{0l} + \gamma_{1l}x$  in (1) yields the linear DDP prior (e.g., DeIorio et al., 2009). Here, the  $(\gamma_{0l}, \gamma_{1l})$  are i.i.d., given hyperparameters, from a centering distribution, typically, with independent components. Hence, the linear DDP reduces the centering stochastic process  $G_{0,\mathcal{X}}$  in (1) to a linear function over dose levels with random component-specific intercept and slope parameters. Now, with the DP truncation approximation, the linear DDP Binomial mixture is given by

$$f(y \mid m; G_x^L) = \sum_{l=1}^L p_l \text{Bin}(y; m, \pi(\gamma_{0l} + \gamma_{1l}x)).$$

It is straightforward to verify that, if  $\gamma_{1l} > 0$  for all  $l$ , then the corresponding dose-response curve is non-decreasing in  $x$ . We implement this model assuming  $\gamma_{0l} \mid \delta_0, \sigma_0^2$  i.i.d.  $N(\delta_0, \sigma_0^2)$ , and, independently,  $\gamma_{1l} \mid \varphi$  i.i.d.  $\text{gamma}(c, \varphi)$ . We fix  $c = 1$  and assign hyperpriors to  $\delta_0$ ,  $\sigma_0^2$  and  $\varphi$ , specifically, a normal prior to  $\delta_0$  with mean 0 and variance 20, an inverse gamma prior to  $\sigma_0^2$  with shape parameter 2 and mean 25, and a  $\text{gamma}(10, 1)$  prior to  $\varphi$ .

### 3.3 Results

Under the first simulation case, Figure 3 gives the posterior mean and 90% uncertainty bands for the probability mass function of the number of negative outcomes given  $m = 12$  implants for three dose levels. The GP and DDP mixture models provide similar inference, with slightly larger uncertainty bands arising from the DDP model. The linear DDP model produces somewhat less accurate point estimates with narrow interval estimates. While the PMDP model captures the general shape of the mass function, there is substantial uncertainty in its interval estimates.

The mixture of Binomials simulation setting provides more striking differences between the performance of the models, as seen in Figure 4. The GP model relies on a Binomial response distribution and therefore can not pick up the bimodality at dose levels 30 and 45. The linear DDP model attempts to capture the essence of the bimodal shapes of the probability mass functions at doses 30 and 45; however its restrictive dependence structure limits the posterior accuracy. Inference at dose level 75 resembles the actual probability mass function, yet fails to include the true values within its narrow uncertainty bands. The PMDP model generally includes the true probabilities within the large uncertainty bands it produces for all three dose levels. However, the changes in the estimates across and within dose levels are quite drastic. In the case of dose level 75, there are 8 data points associated with  $m = 12$  implants (compared to 15-20 observations for doses 30 and 45). Of these 8 observations, two animals had 5 negative outcomes, which is apparent in the PMDP model results. Owing to the smoother evolution of DDP realizations and to its mixture structure, the DDP model has the capacity to avoid such sudden changes in the estimated probabilities. Moreover, the DDP model recovers the true

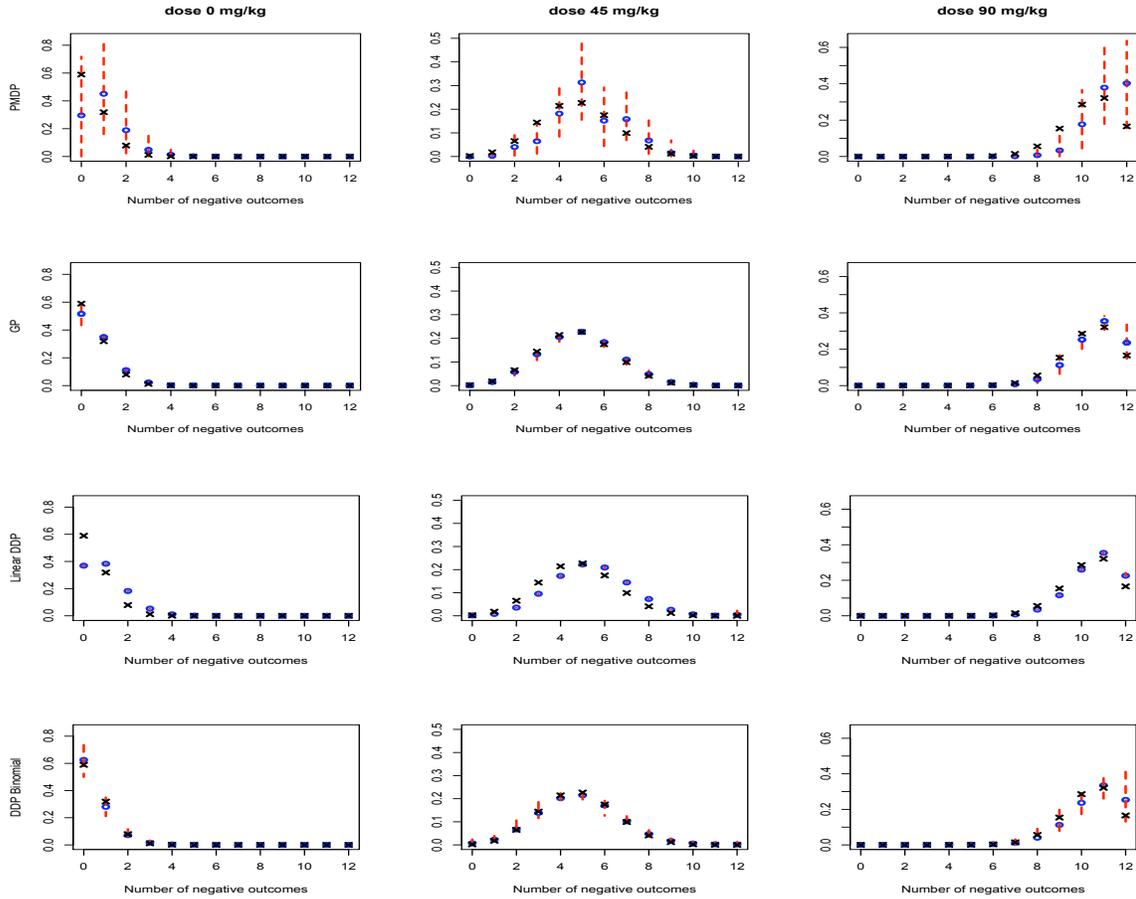


Figure 3: Simulation case 1. Posterior mean (denoted by “o”) and 90% uncertainty bands for the probability mass function of the number of negative outcomes, given  $m = 12$  implants, at three dose levels, using the PMDP, GP, Linear DDP, and DDP models (from top to bottom row). In each panel, the values of the true probability mass function are denoted by “x”. The online version of this figure is in color.

probability mass function shapes with notably tighter uncertainty bands than the PMDP model.

Contrasting the results from Figures 3 and 4 reveals an appealing feature of the proposed modeling framework: the DDP mixture model can uncover non-standard distributional shapes at different dose levels when such shapes are suggested by the data (Figure 4), but at the same time, will recover simpler probability mass functions with a relatively small amount of additional uncertainty relative to parametric models (Figure 3).

Because inference for the entire dose-response curve is not readily available under the PMDP

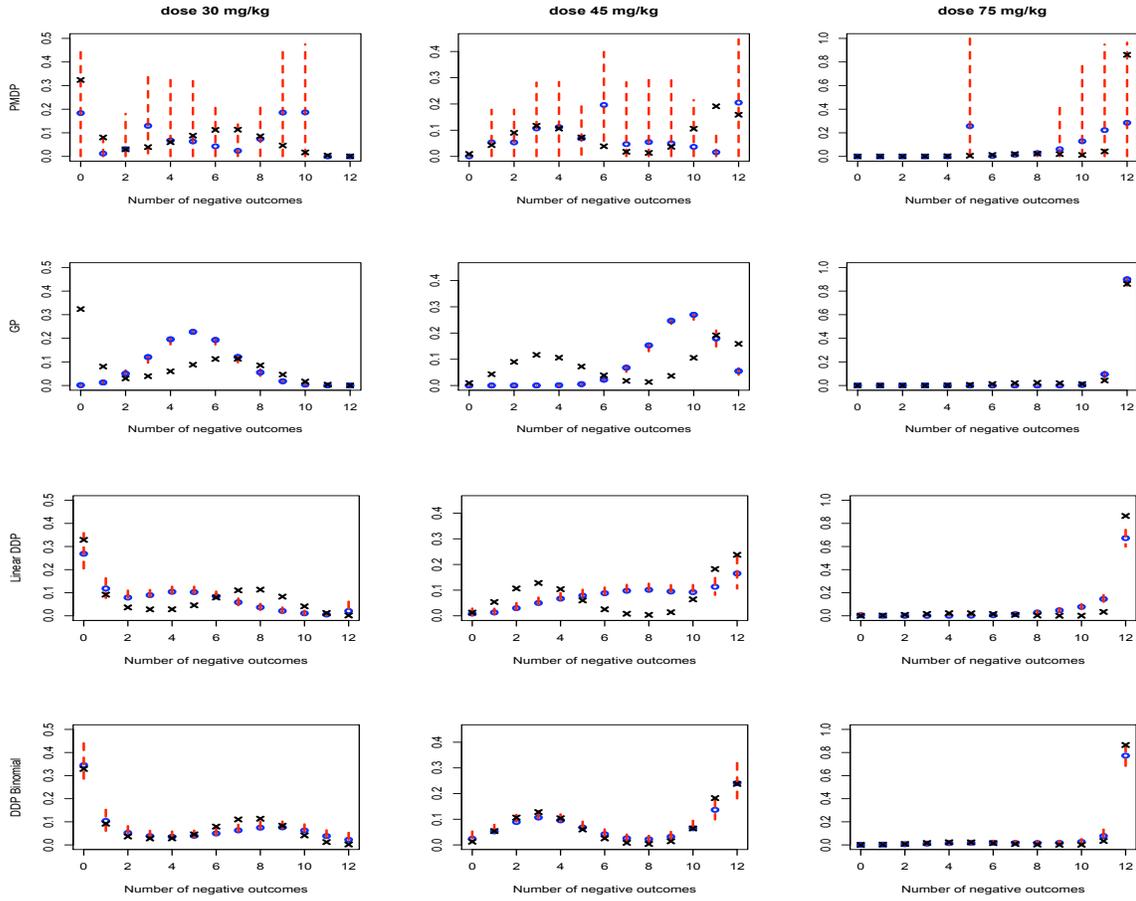


Figure 4: Simulation case 2. Posterior mean (denoted by “o”) and 90% uncertainty bands for the probability mass function of the number of negative outcomes, given  $m = 12$  implants, at three dose levels, using the PMDP, GP, Linear DDP, and DDP models (from top to bottom row). In each panel, the values of the true probability mass function are denoted by “x”. The online version of this figure is in color.

model, we focus this aspect of the comparison on the GP, linear DDP, and DDP mixture models. For the first simulation setting, the DDP model represents more accurately the dose-response curve across dose levels (Figure 5). While the GP model performs well at the observed dose levels, it fares worse at the interpolated doses as indicated by the jumps at the observed doses in the posterior mean dose-response curve. Under the linear DDP model, dependence across dose  $x$  is built solely through the linear functions  $\gamma_{0l} + \gamma_{1l}x$ , and we thus expect roughly uniform uncertainty in posterior inference results for the dose-response curve. This is reflected in Figure

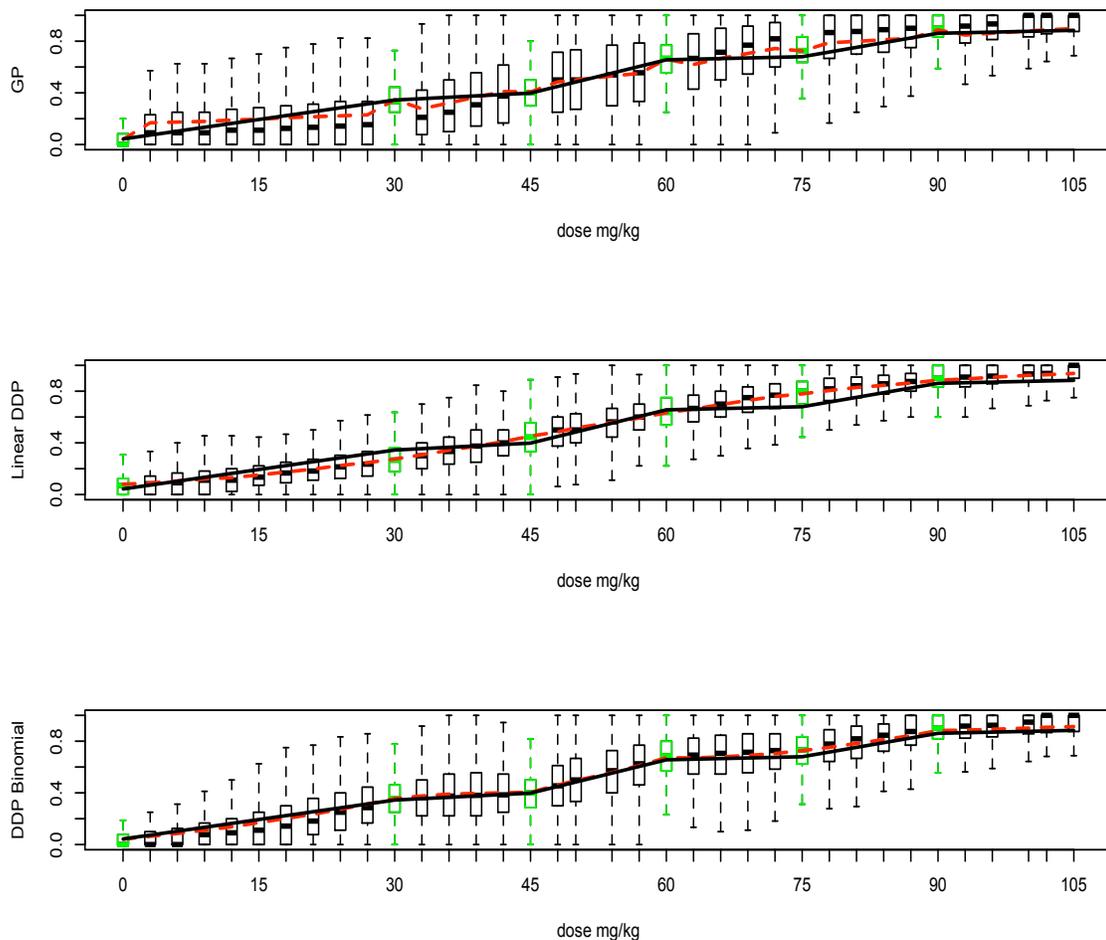


Figure 5: Simulation case 1. Boxplots of the posterior predictive samples of  $y_0/m_0$  at the observed and new dose levels, using the GP model (top), linear DDP (middle), and DDP mixture model (bottom). In each panel, the posterior mean estimate for the dose-response curve and the true curve are denoted by the dashed and solid line, respectively. The online version of this figure is in color.

5, where we also note that the linear DDP posterior mean estimate tends to smooth out the probability of response. This smoothing results in biased probability mass function estimates, as discussed previously. For the second simulation example, the DDP model again produced a more accurate estimate for the dose-response function (results not shown), although in this case with larger posterior predictive uncertainty compared to the GP and linear DDP models.

Table 1: Simulation study. Posterior mean and 0.05 and 0.95 percentiles (in parentheses) for the probability of a negative outcome at the 6 observed dose levels, using the GP, linear DDP, DDP mixture, and PMDP models. The true values of the dose-response curve are given in bold.

Simulation case 1					
Dose	GP	Linear DDP	DDP	PMDP	
0 ( <b>0.04</b> )	0.04 (0.03,0.05)	0.08 (0.07,0.09)	0.04 (0.03,0.06)	0.09 (0.00,0.06)	
30 ( <b>0.34</b> )	0.35 (0.32,0.38)	0.28 (0.27,0.29)	0.36 (0.33,0.39)	0.30 (0.03,0.72)	
45 ( <b>0.39</b> )	0.41 (0.38,0.43)	0.45 (0.44,0.46)	0.40 (0.37, 0.43)	0.47 (0.08,0.89)	
60 ( <b>0.66</b> )	0.67 (0.63,0.70)	0.63 (0.62,0.64)	0.67 (0.62,0.70)	0.64 (0.24,0.95)	
75 ( <b>0.68</b> )	0.73 (0.69,0.76)	0.78 (0.77,0.79)	0.73 (0.69,0.76)	0.79 (0.34,0.99)	
90 ( <b>0.86</b> )	0.89 (0.86,0.92)	0.88 (0.87,0.89)	0.88 (0.83,0.91)	0.88 (0.45,1.00)	
Simulation case 2					
Dose	GP	Linear DDP	DDP	PMDP	
0 ( <b>0.02</b> )	0.03 (0.02,0.04)	0.04 (0.03,0.06)	0.03 (0.02,0.05)	0.09 (0.00,0.07)	
30 ( <b>0.34</b> )	0.34 (0.31,0.36)	0.27 (0.23,0.32)	0.33 (0.28,0.39)	0.34 (0.01,0.89)	
45 ( <b>0.60</b> )	0.61 (0.59,0.63)	0.65 (0.62,0.69)	0.60 (0.55,0.66)	0.58 (0.06,0.98)	
60 ( <b>0.88</b> )	0.87 (0.86,0.89)	0.87 (0.85,0.90)	0.87 (0.83,0.91)	0.79 (0.09,1.00)	
75 ( <b>0.94</b> )	0.93 (0.91,0.95)	0.94 (0.92,0.95)	0.92 (0.88,0.96)	0.89 (0.28,1.00)	
90 ( <b>0.95</b> )	0.97 (0.96,0.98)	0.97 (0.95,0.98)	0.96 (0.92,0.99)	0.97 (0.85,1.00)	

Finally, we compare the four models with regard to inference for the probability of a negative outcome at the six observed dose levels  $x_i$ . Denoting as before by  $y^*$  a generic binary outcome at  $x_i$ , this probability is given by  $\sum_{l=1}^L p_l \pi(Z_l(x_i))$  under the DDP model,  $\sum_{l=1}^L p_l \pi(\gamma_{0l} + \gamma_{1l} x_i)$  under the linear DDP model, and  $\pi(\theta(x_i))$  under the GP model; moreover, under the PMDP model, it arises from a  $\text{Beta}(A_i \pi(\eta_0 + \eta_1 x_i), A_i (1 - \pi(\eta_0 + \eta_1 x_i)))$  distribution. Table 1 includes point and 90% interval estimates based on the corresponding posterior distributions. Noteworthy here are the results under the PMDP model, which produces interval estimates that are too wide to be practical. This level of posterior uncertainty is consistent with the results in Figures 3 and 4. For the 2,4,5-T data considered in Section 4.1, we also obtained overly wide 90% interval estimates from the PMDP model (with the same interquartile ranges in Table 1 of Dominici and Parmigiani, 2001). Contrarily, for the first simulation case, the linear DDP model moderates the irregularities of the actual dose-response function into a smooth curve (see Figure 5), resulting in underestimation of the variability in the negative outcome probability. The model produces more

realistic interval estimates in the second case, though it generally overestimates the probabilities. As suggested by the posterior predictive results for  $y_0/m_0$ , the GP and DDP mixture models yield relatively similar inference for the dose-response curve at the observed dose levels. The DDP model produces wider interval estimates, especially under the second simulation setting, which are more effective in capturing the true values at the largest dose.

## 4 Data illustrations

We illustrate the proposed nonparametric modeling approach with the 2,4,5-T data (Section 4.1) and the DEHP data (Section 4.2), which were introduced in Section 1.2.

### 4.1 Application to 2,4,5-T data

Focusing first on inference for conditional response distributions, Figure 6 plots the posterior mean and 90% uncertainty bands for  $f(y \mid m = 12; G_{x_i}^L)$  at all observed dose levels, and for  $f(y \mid m = 12; G_{\tilde{x}_0}^L)$  at two new doses, one ( $\tilde{x}_0 = 50$  mg/kg) within the observed range, and one extrapolated at  $\tilde{x}_0 = 100$  mg/kg. The probability mass functions corresponding to low and high dose levels depict shapes that could be captured by traditional parametric models. However, in the mid-range of dose values, the DDP mixture model uncovers non-standard probability mass function shapes, which suggest bimodality. The estimated mass functions at the new dose levels have larger probability bands, and their shape highlights the smooth evolution of the DDP-based response distributions across dose levels.

The posterior densities for the intracluster correlations at the observed dose levels are given in the left panel of Figure 7. The correlations depict an increasing trend up to dose levels 60–75 mg/kg, with increasing uncertainty beyond dose 75 mg/kg consistent with the smaller number of dams at the two higher dose levels. Using for illustration four probabilities,  $q = 0.1, 0.25, 0.4$ , and  $0.5$ , the right panel of Figure 7 shows the posterior densities of the corresponding calibrated dose level  $x_q$ , obtained as discussed in Section 2.4.2.

While results for response distributions are not shown here, we also fit the Binomial-logit and

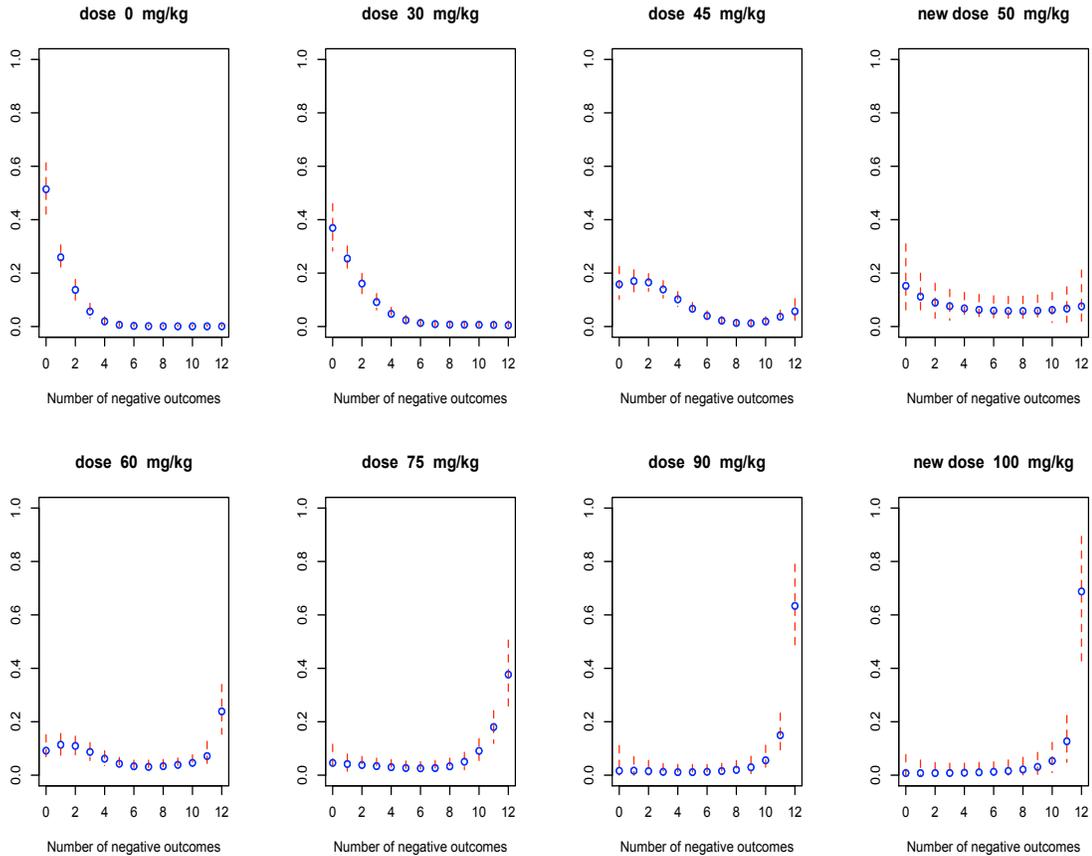


Figure 6: 2,4,5-T data. For the 6 observed dose levels and 2 new dose levels, the posterior mean probability mass functions (denoted by “o”) and 90% uncertainty bands for the number of negative outcomes conditional on  $m = 12$  implants. The online version of this figure is in color.

Beta-Binomial models to the 2,4,5-T data. The Binomial-logit model is not flexible enough to capture the non-standard distributions and estimates little variation. The Beta-Binomial model also cannot deviate from unimodal probability mass functions, but attempts to compensate for the data heterogeneity by increasing the variability in the probability of response, thereby producing large uncertainty bands. This overcompensation is manifested in the overly wide interval estimates for the dose-response curve under the Beta-Binomial model (Figure 8, middle panel). On the other extreme, the Binomial-logit model underestimates the uncertainty in the curve, and is also restricted to the logistic function shape (Figure 8, left panel). The posterior mean estimate from the DDP mixture model (Figure 8, right panel) supports a non-decreasing

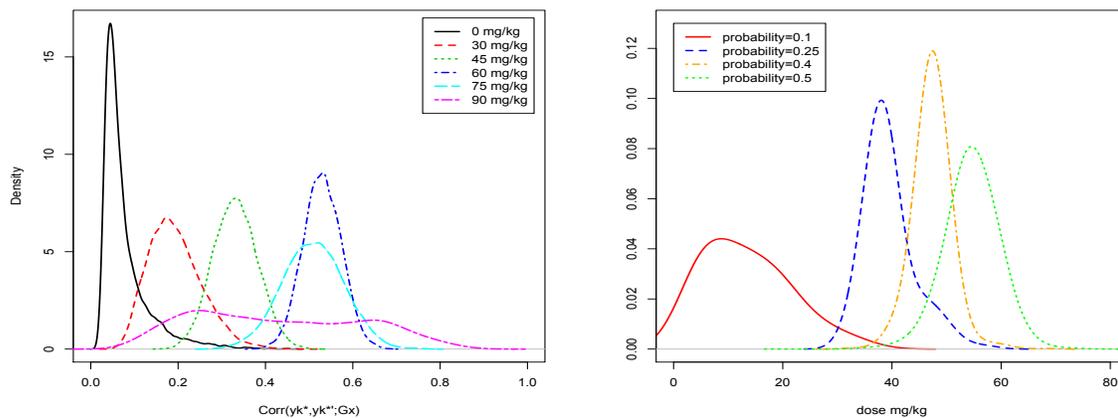


Figure 7: 2,4,5-T data. The left panel plots the posterior densities for the intracluster correlations at each of the six observed dose levels. The posterior densities for the calibrated dose level corresponding to four probability thresholds are given in the right panel. The online version of this figure is in color.

dose-response relationship with curvature that deviates at smaller doses from the logistic shape, and with larger uncertainty at the interpolated values.

Widening uncertainty bands in interpolating/extrapolating regions is a general characteristic of inference under nonparametric prior models. In our setting, it is intensified by the fact that developmental toxicology data typically comprise a larger number of dams at the lower dose levels. Although results are not reported in Section 3, we have also experimented with simulated data involving the same number of dams per dose level. With balanced response replicates, we again obtained widening uncertainty bands in the interpolating regions for the dose-response curve, but the amount of narrowing was comparable at all observed dose levels.

The inference results for the 2,4,5-T dose-response curve provide an interesting illustration of a nonparametric Bayesian model producing more realistic uncertainty quantification for posterior estimates relative to simpler parametric models. In particular, in contrast to continuous mixing that defines the Beta-Binomial model, the discrete nature of the DDP prior enables clustering of the Binomial kernel latent mixing parameters, thus, controlling more effectively the variability of the estimated conditional response distributions.

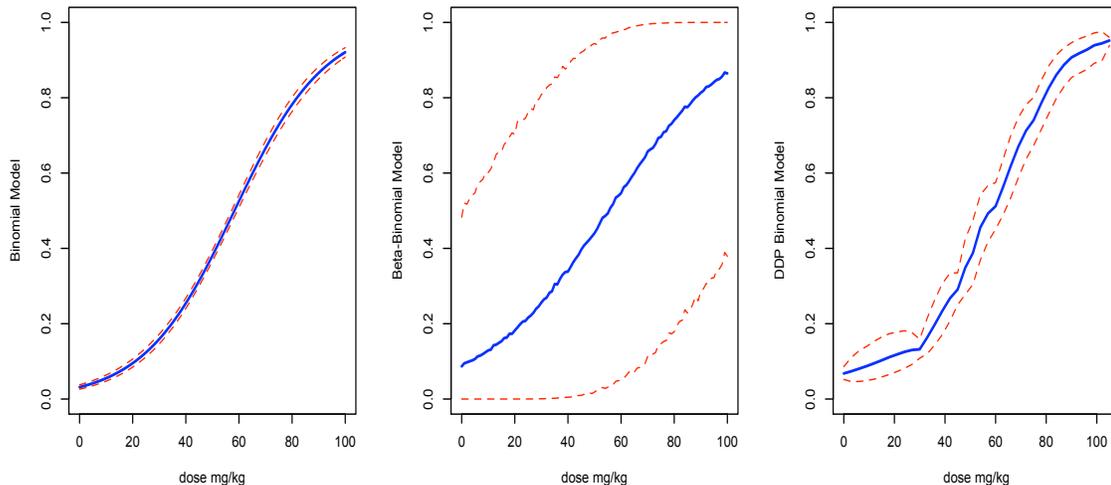


Figure 8: 2,4,5-T data. The posterior mean estimate (solid line) and 90% uncertainty bands (dashed lines) for the dose-response curve, using the Binomial-logit model (left panel), the Beta-Binomial model (middle panel), and the DDP Binomial mixture model (right panel). The online version of this figure is in color.

## 4.2 Application to DEHP data

Here, we present a brief analysis of the DEHP data, mainly to highlight the feature of the DDP modeling framework with regard to recovering non-monotonic dose-response relationships.

First, we note that the data (Figure 9, left panel) appear to suggest a drop in the probability of a negative outcome from the control level to level 25 mg/kg  $\times$  1000. In addition to the graphical indication, the drop is suggested by an (admittedly crude) “data-based” analysis, using independent Binomials with common probability for all dams at each dose. The resulting (maximum likelihood) estimates of the probability of a negative outcome at doses 0 and 25mg/kg  $\times$  1000 are equal to 0.200 and 0.116, with respective standard errors 0.0209 and 0.0179. As discussed in Section 1.2, such a dose-response pattern may be associated with hormesis, and thus, it is practically important to be able to quantify how well it is supported by the data.

Indeed, this particular non-monotonic dose-response shape is apparent in the posterior mean estimate and corresponding uncertainty bands for  $\Pr(y^* = 1; G_{\chi}^L)$  (Figure 9, right panel). Under

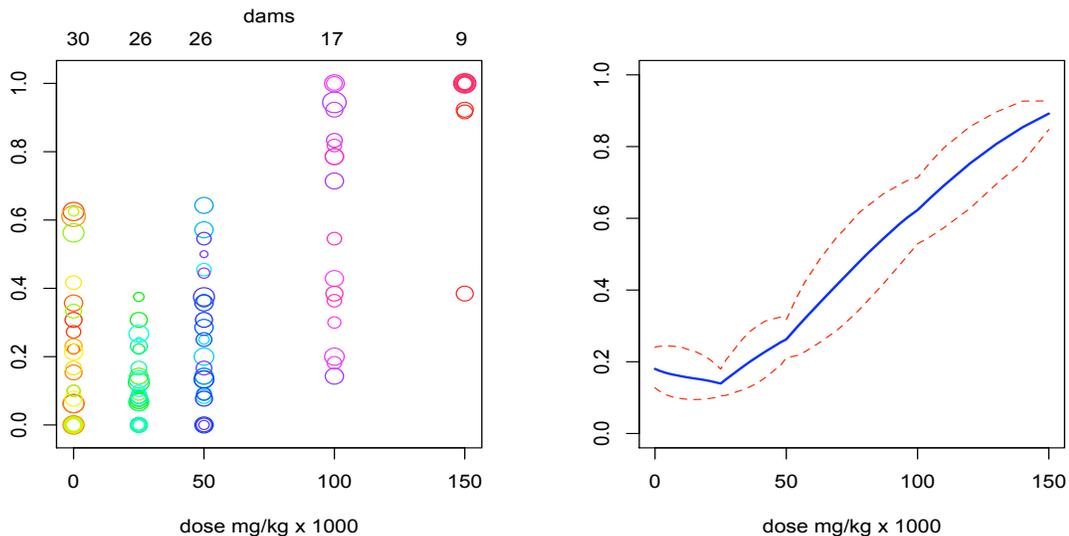


Figure 9: DEHP data. The left panel shows the data, where each circle corresponds to a dam, the size of the circle is proportional to the number of implants, and the coordinates of the circle are the dose level and the proportion of negative outcomes. The right panel includes the posterior mean estimate (solid line) and 90% uncertainty bands (dashed lines) for the dose-response curve. The online version of this figure is in color.

essentially all standard models for developmental toxicology data, this dip in the dose-response curve would not be captured. Moreover, the DDP Binomial mixture model was again able to recover varying shapes for conditional response distributions across dose levels. In particular, point and interval estimates for  $f(y | m = 12; G_x^L)$  (not shown) support shapes that evolve with increasing dose from right to left skewness, with bimodal probability mass functions uncovered for values of  $x$  around observed dose 100 mg/kg  $\times$  1000.

## 5 Discussion

We have developed a Bayesian nonparametric mixture framework for modeling and risk assessment in developmental toxicity studies. The impetus for the proposed modeling approach is that for such studies it is critical to model flexibly both the dam specific distributions and the probability of response to accurately account for the multiple sources of data heterogeneity. The methodology is built from Binomial mixtures with a dependent Dirichlet process prior for the

dose-dependent mixing distributions. The resulting nonparametric mixture model provides rich inference for the response distribution as well as for the dose-response curve. Using a simulation study, we have shown that, relative to simpler semiparametric Bayesian approaches, the DDP mixture model is the only one that accomplishes both of the inferential goals above. Finally, data from two toxicity studies were used to illustrate the variety of inferences that can be obtained from the DDP mixture model, including its practical utility with regard to estimation of non-monotonic dose-response relationships.

Note that in our nonparametric mixture model formulation, the choice of the single- $p$  DDP prior strikes a good balance between model flexibility and computational feasibility. Data from traditional toxicology experiments have on the order of 5-10 dose levels and are not likely to indicate drastic changes in distributional shapes between nearby dose levels. While the PMDP model is heavily influenced by individual data points (e.g., Figure 4), the single- $p$  DDP mixture prior induces a smooth evolution across the dose levels. The small number of dose levels may be problematic for learning about the parameters in more general models where the DDP prior weights are also dose dependent. As the data examples have demonstrated, the single- $p$  DDP prior mixture model is sufficiently flexible to capture the dependence structure of the distributions across dose levels, while remaining interpretable and manageable to implement.

Further simplification of the single- $p$  DDP mixture model, based on the linear DDP prior as developed in Section 3.2, sacrifices desirable flexibility in the context of developmental toxicology data analysis. Although MCMC computing is simplified relative to the general DDP mixture model, the restrictive functional dependence on the dose level may be too limited to capture the complexity of the data, and the implied non-decreasing dose-response curve may be inappropriate in some experiments (see Section 4.2). However, the linear DDP mixture approach is promising for more traditional bioassay settings for which there is no nested structure in the data at each dose level, and where the monotonicity assumption for the dose-response curve is more universally accepted; results in this direction are reported in Fronczyk and Kottas (2012).

A practically important extension involves modeling developmental toxicology data with

responses that include the multicategory classification (“dead”, “normal”, “malformed”) as in the more general data structure discussed in Section 1.2. Now, the mixture model for a generic dam with  $m$  implants at dose level  $x$  can be built from kernel  $\text{Bin}(R; m, \pi(\gamma))\text{Bin}(y'; m - R, \pi(\theta))$ , with DDP mixing on both  $\gamma$  and  $\theta$ , where  $R$  is the number of non-viable fetuses (including resorptions and prenatal deaths) and  $y'$  is the number of malformations. Under this modeling approach, of interest is inference for risk assessment associated with both the probability of “death” and the probability of “malformation”. The DDP mixture modeling framework can be further extended to include continuous responses (say, fetal weight) for each pup. Here, the equivalent formulation in (3) for the DDP Binomial mixture is central as it enables modeling for the data at the pup level. Key inferential objectives include study of the effect of the exposure level to the toxin on fetal malformations, prenatal death rates, and fetal weight at term.

## Appendix: Properties of the DDP mixture model

**Monotonicity of the prior expectation for the dose-response curve:** Denote by  $D(x)$ ,  $x \in \mathcal{X}$ , the prior expectation for the dose-response curve. As discussed in Section 2.3, there is no explicit assumption of monotonicity for the dose-response curve under the DDP mixture model. However,  $D(x)$  is a non-decreasing function provided  $\beta_1 > 0$ .

Under the DDP truncation approximation,  $D(x) = \text{E}(\text{Pr}(y^* = 1; G_x^L)) = \sum_{l=1}^L \text{E}(p_l)\text{E}(\pi(Z_l(x)))$ , since  $\{Z_l(x) : l = 1, \dots, L\}$  is independent of  $\{V_l : l = 1, \dots, L-1\}$ , the collection of i.i.d.  $\text{Beta}(1, \alpha)$  variables that define the  $p_l$  through stick-breaking. Therefore, for any  $x < x'$ ,  $D(x) - D(x') = \sum_{l=1}^L \text{E}(p_l)\{\text{E}(\pi(Z_l(x))) - \text{E}(\pi(Z_l(x')))\}$ . Now, for any  $l = 1, \dots, L$ ,  $Z_l(x)$  and  $Z_l(x')$  follow  $\text{N}(\beta_0 + \beta_1 x, \sigma^2)$  and  $\text{N}(\beta_0 + \beta_1 x', \sigma^2)$  distributions, respectively. Hence, if  $\beta_1 > 0$ ,  $Z_l(x)$  is stochastically smaller than  $Z_l(x')$ , for each  $l = 1, \dots, L$ . This in turn implies  $\text{E}(\pi(Z_l(x))) \leq \text{E}(\pi(Z_l(x')))$ , for all  $l = 1, \dots, L$  (since  $\pi(u)$  is an increasing function), and thus  $D(x) \leq D(x')$ .

The above argument establishes the result for the form in (5), which is the one we work with to obtain inference for the dose-response relationship. The result can also be obtained

without the truncation approximation. In this case, we have  $D(x) = E(\Pr(y^* = 1; G_x)) = E\{\int \pi(\theta)dG_x(\theta)\} = \int \pi(\theta)dG_{0x}(\theta)$ , where  $G_{0x} = N(\beta_0 + \beta_1 x, \sigma^2)$ . Therefore,  $D(x)$  is the expectation of the (increasing) logistic function with respect to  $G_{0x}$ , which is stochastically ordered in  $x$  provided  $\beta_1 > 0$ , and thus  $D(x)$  is a non-decreasing function of  $x$ .

**Positive intracluster correlation:** Consider the vector of binary responses,  $\mathbf{y}^* = (y_1^*, \dots, y_m^*)$ , for a generic dam with  $m$  ( $\geq 2$ ) implants at dose level  $x$ . Let  $\pi^* = \Pr(y_k^* = 1; G_x) = \int \pi(\theta)dG_x(\theta)$ , and denote by  $\gamma = \text{Corr}(y_k^*, y_{k'}^*; G_x)$  the correlation between any pair of binary outcomes within the same dam;  $\gamma$  is given by (4) under the DDP truncation approximation.

Under the implicit assumption of common  $\pi^*$  and  $\gamma$  for all binary responses within the same dam, the variance for the number of combined negative outcomes,  $y = \sum_{k=1}^m y_k^*$ , is given by  $\text{Var}(y | m; G_x) = m\pi^*(1 - \pi^*)\{1 + (m - 1)\gamma\}$ . (Note that this result does not rely on the specific form of the mixture model for  $y$  in (2) or the equivalent model for  $\mathbf{y}^*$  in (3).) Now, consider a random variable  $u$ , which has a Binomial distribution with the same mean as  $y$  arising from  $f(y | m; G_x) = \int \text{Bin}(y; m, \pi(\theta))dG_x(\theta)$ , that is,  $u \sim \text{Bin}(m, \pi^*)$ . Then, using overdispersion results for mixtures from exponential families (e.g., Shaked, 1980), we have  $\text{Var}(y | m; G_x) \geq \text{Var}(u) = m\pi^*(1 - \pi^*)$ , which yields  $\gamma \geq 0$ .

## References

- Antoniak, C. (1974), "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *The Annals of Statistics*, 2, 1152–1174.
- Berry, D. A. and Christensen, R. (1979), "Empirical Bayes estimation of a Binomial parameter via mixtures of Dirichlet processes," *The Annals of Statistics*, 7, 558–568.
- Bowman, D. and George, E. O. (1995), "A saturated model for analyzing exchangeable binary data: applications to clinical and developmental toxicity studies," *Journal of the American Statistical Association*, 90, 871–879.
- Calabrese, E. J. (2005), "Paradigm lost, paradigm found: The re-emergence of hormesis as a fundamental dose response model in the toxicological sciences," *Environmental Pollution*, 138, 378–411.

- Carota, C. and Parmigiani, G. (2002), “Semiparametric regression for count data,” *Biometrika*, 89, 265–281.
- Catalano, P. J. and Ryan, L. M. (1992), “Bivariate latent variable models for clustered discrete and continuous outcomes,” *Journal of the American Statistical Association*, 87, 651–658.
- Chen, J. J., Kodell, R. L., Howe, R. B., and Gaylor, D. W. (1991), “Analysis of trinomial responses from reproductive and developmental toxicity experiments,” *Biometrics*, 47, 1049–1058.
- DeIorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009), “Bayesian nonparametric nonproportional hazards survival modeling,” *Biometrics*, 65, 762–771.
- DeIorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004), “An ANOVA model for dependent random measures,” *Journal of the American Statistical Association*, 99, 205–215.
- Dominici, F. and Parmigiani, G. (2001), “Bayesian semiparametric analysis of developmental toxicology data,” *Biometrics*, 57, 150–157.
- Dunson, D., Chen, Z., and Harry, J. (2003), “A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes,” *Biometrics*, 59, 521–530.
- Escobar, M. and West, M. (1995), “Bayesian density estimation and inference using mixtures,” *Journal of the American Statistical Association*, 90, 577–588.
- Faes, C., Geys, H., Aerts, M., and Molenberghs, G. (2006), “A hierarchical modeling approach for risk assessment in developmental toxicity studies,” *Computational Statistics & Data Analysis*, 51, 1848–1861.
- Ferguson, T. (1973), “A Bayesian analysis of some nonparametric problems,” *The Annals of Statistics*, 1, 209–230.
- Fronczyk, K. and Kottas, A. (2012), “A Bayesian approach to the analysis of quantal bioassay studies using nonparametric mixture models,” Technical Report UCSC-SOE-12-01, Department of Applied Mathematics and Statistics, University of California, Santa Cruz.
- Gelfand, A. E., Kottas, A. and MacEachern, S. N. (2005), “Bayesian nonparametric spatial modeling with Dirichlet process mixing,” *Journal of the American Statistical Association*, 100, 1021–1035.
- George, E. O. and Bowman, D. (1995), “A full likelihood procedure for analysing exchangeable binary data,” *Biometrics*, 51, 512–523.
- Holson, J. F., Gaines, T. B., Nelson, C. J., LaBorde, J. B., Gaylor, D. W., D.M. Sheehan, D. M., and Young, J. F. (1991), “Developmental toxicity of 2,4,5-trichlorophenoxyacetic acid I: Dose response studies in four inbred strains and one outbred stock of mice,” *Fundamentals of Applied Toxicology*, 19, 286–297.

- Ishwaran, H. and James, L. (2001), “Gibbs sampling methods for stick-breaking priors,” *Journal of the American Statistical Association*, 96, 161–173.
- Kottas, A., Duan, J., and Gelfand, A. E. (2008), “Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures,” *Biometrical Journal*, 50, 29–42.
- Kottas, A. and Krnjajić, M. (2009), “Bayesian semiparametric modelling in quantile regression,” *Scandinavian Journal of Statistics*, 36, 297–319.
- Kuk, A. Y. C. (2004), “A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions,” *Applied Statistics*, 53, 369–386.
- Liu, J. S. (1996), “Nonparametric hierarchical Bayes via sequential imputations,” *The Annals of Statistics*, 24, 911–930.
- MacEachern, S. N. (2000), “Dependent Dirichlet processes,” Technical Report, Department of Statistics, The Ohio State University.
- Molenberghs, G. and Ryan, L. M. (1999), “An exponential family model for clustered multivariate binary data,” *Environmetrics*, 10, 279–300.
- Pang, Z. and Kuk, A. Y. C. (2005), “A shared response model for clustered binary data in developmental toxicity studies,” *Biometrics*, 61, 1076–1084.
- Regan, M. M. and Catalano, P. J. (1999), “Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology,” *Biometrics*, 55, 760–768.
- Rodriguez, A. and ter Horst, E. (2008), “Bayesian dynamic density estimation,” *Bayesian Analysis*, 3, 339–366.
- Ryan, L. (1992), “Quantitative risk assessment for developmental toxicity,” *Biometrics*, 48, 163–174.
- Sethuraman, J. (1994), “A constructive definition of Dirichlet priors,” *Statistica Sinica*, 4, 639–650.
- Shaked, M. (1980), “On mixtures of exponential families,” *Journal of the Royal Statistical Society, Series B*, 42, 192–198.
- Tyl, R. W., Jones-Price, C., Marr, M. C., and Kimmel, C. A. (1983), “Teratologic evaluation of diethylhexyl phthalate (CAS No. 111-81-7),” Final Study Report for NCTR/NTP contract 222-80-2031(c), NITS PB85105674, National Technical Information Service, Springfield, Virginia.
- Zhang, T. and Liu, J. S. (2012), “Nonparametric hierarchical Bayes analysis of binomial data via Bernstein polynomial priors,” *The Canadian Journal of Statistics*, 40, 328–344.
- Zhu, Y., Krewski, D., and Ross, W. H. (1994), “Dose-response models for correlated multinomial data from developmental toxicity studies,” *Applied Statistics*, 43, 583–598.