Risk assessment for toxicity experiments with discrete and continuous outcomes: A Bayesian nonparametric approach

Kassandra Fronczyk and Athanasios Kottas *

Abstract: We present a Bayesian nonparametric modeling approach to inference and risk assessment for developmental toxicity studies. The primary objective of these studies is to determine the relationship between the level of exposure to a toxic chemical and the probability of a physiological or biochemical response. We consider a general data setting involving clustered categorical responses on the number of prenatal deaths, the number of live pups, and the number of live malformed pups from each laboratory animal, as well as continuous outcomes (e.g., body weight) on each of the live pups. We utilize mixture modeling to provide flexibility in the functional form of both the multivariate response distribution and the various dose-response curves of interest. The nonparametric model is built from a structured mixture kernel and a dose-dependent Dirichlet process prior for the mixing distribution. The modeling framework enables general inference for the implied dose-response relationships and for dose-dependent correlations between the different endpoints, features which provide practical advances relative to traditional parametric models for developmental toxicology. We use data from a toxicity

*K. Fronczyk, Applied Statistics Group, Lawrence Livermore National Laboratory, Livermore, CA, USA. (E-mail: *fornczyk1@llnl.gov*). A. Kottas, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, USA. (E-mail: *thanos@soe.ucsc.edu*).

experiment that investigated the toxic effects of an organic solvent (diethylene glycol dimethyl ether) to demonstrate the range of inferences obtained from the nonparametric mixture model, including comparison with a parametric hierarchical model.

KEY WORDS: Dependent Dirichlet process; Developmental toxicology data; Dose-response relationship; Gaussian process; Nonparametric mixture modeling.

1 Introduction

Developmental toxicity studies, a generalization of the standard bioassay setting, investigate birth defects induced by toxic chemicals. The most common type of developmental toxicology data structure arises from the Segment II design, where at each experimental dose level, a number of laboratory animals (or dams) are exposed to the toxin after implantation. Recorded from each dam are the number of implants, the number of resorptions (i.e., undeveloped embryos or very early fetal deaths) and prenatal deaths, the number of live pups, and the number of live malformed pups. Resorptions and prenatal deaths are typically combined in the available data sets, and we interchangeably refer to this endpoint as non-viable fetuses or prenatal deaths. Additional outcomes measured on each of the live pups may include body weight and length.

The main objective of this type of toxicity studies is to examine the relationship between the level of exposure to the toxin (dose level) and the different endpoints, which include prenatal death (embryolethality), malformation, and low birth weight. The dose-response curve for each endpoint is defined by the probability of the corresponding outcome across 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. There are a number of quantities examined for risk assessment, including conditional prob-

abilities of an outcome given specific conditions and correlations between responses.

Incorporating into the modeling approach a continuous outcome, such as weight at birth, for each of the live pups presents a challenge in that there are now clustered outcomes that include both discrete and continuous responses. The related literature includes a plethora of likelihood based estimation methods (e.g., Catalano & Ryan, 1992; Regan & Catalano, 1999; Gueorguieva & Agresti, 2001); however, these approaches rely on restrictive parametric assumptions and are limited with regard to uncertainty quantification for risk assessment. Regarding Bayesian work, we are aware of only two parametric approaches. Dunson et al. (2003) propose a joint model for the number of viable fetuses and multiple discrete-continuous outcomes. A continuation-ratio ordinal response model is used for the number of viable fetuses and the multiple outcomes are assigned an underlying normal model with shared latent variables within outcome-specific regression models. In Faes et al. (2006), the proposed model is expressed in two stages; the first models the probability that a fetus is non-viable, and the second determines the probability that a viable fetus has a malformation and/or suffers from low birth weight.

For illustrative purposes, we will focus on a study – available from the National Toxicology Program database – where diethylene glycol dimethyl ether (DYME), an organic solvent, is evaluated for toxic effects in pregnant mice (Price et al., 1987). This data example (see Figure 1) includes a small set of four active dose levels, with a comparable number of animals exposed at each dose level ($18 - 24$ dams). The variability in the discrete responses is vast, due to the inherent heterogeneity of both the dams and the pups' reaction to the toxin. For both endpoints of embryolethality and malformation, an increasing trend across toxin levels is suggested, although with no obvious parametric choices for the associated dose-response curves. Large variability is also evident in the birth weight responses for which a decreasing dose-response relationship is indicated. This

complex nature of the DYME data is representative of the data structures that arise from developmental toxicity experiments. Hence, the modeling approach needs to account for the multiple sources of variability and simultaneously relax potentially restrictive assumptions on all inferential objectives.

We provide a comprehensive framework built upon nonparametric mixture modeling, which results in flexibility in both the collection of response distributions as well as the form of the dose-response relationship for the multiple clustered endpoints. The dependence of the response distributions is governed by the dose level, implying that distributions corresponding to nearby dose levels are more closely related than those far apart. The mixture model provides a means to quantify the variability in the response distributions, which carries over to the dose-response relationships. The assumptions of the mixture model bestow a foundation for interpolation and extrapolation of the dose-response curves at unobserved dose levels.

The Bayesian nonparametric model developed here extends our earlier work for discrete outcomes (Fronczyk & Kottas, 2014; Kottas & Fronczyk, 2013). As detailed in Section 2, the methodology for the general setting with discrete-continuous outcomes involves non-trivial extensions in the mixture model formulation with respect to properties of the multiple dose-response curves, as well as in the Markov chain Monte Carlo (MCMC) posterior simulation method. To our knowledge, the literature does not include other Bayesian nonparametric methods for developmental toxicology data with a multicategory response classification or with clustered discrete-continuous responses. A Bayesian semiparametric model for the univariate case of combined prenatal death and malformation endpoints was proposed in Dominici & Parmigiani (2001) and further extended by Nott & Kuk (2009). Hwang & Pennell (2013) develop a semiparametric prior model for binary and continuous responses, which is however applicable only to dam-level outcomes, that is, it does not

incorporate the clustered binary and continuous pup-level outcomes.

The paper continues as follows. In Section 2, we develop the nonparametric mixture model and study the dose-response curves for the different endpoints. Section 3 illustrates the range of inferences obtained from the nonparametric model using the DYME data, including comparison with a parametric hierarchical model. Finally, concluding remarks are found in Section 4. The Supplementary Material includes MCMC posterior simulation details, numerical summaries for the DYME data, as well as additional inference results to the ones reported in Section 3.

2 Methods

2.1 Model development

To fix notation, consider a given experimental dose level, x, and a number of pregnant laboratory animals (dams) exposed to the toxin at level x. A generic dam, exposed to dose x, has m implants of which the number of prenatal deaths are recorded as R. Available from the $m-R$ live pups are binary malformation responses, $\boldsymbol{y}^* = \{y_k^* : k = 1, \ldots, m-R\}$, and continuous (birth weight) responses, $\boldsymbol{u}^* = \{u_k^* : k = 1, \ldots, m-R\}$.

Although the number of implants is a random variable, it is natural to assume that its distribution is not dose dependent for Segment II toxicity experiments where exposure occurs after implantation. We thus build the joint probability model for $(m, R, \boldsymbol{y}^*, \boldsymbol{u}^*)$ through $f(m)f(R, \boldsymbol{y}^*, \boldsymbol{u}^* \mid m)$, where only the latter distribution depends on dose level x. Hence, inference for the parameters of the implant distribution is carried out separately from inference for the parameters of the model for $f(R, \boldsymbol{y}^*, \boldsymbol{u}^* \mid m)$. Although more general models can be utilized, we work with a shifted Poisson implant distribution, that is, $f(m) \equiv f(m \mid \lambda) = e^{-\lambda}\lambda^{m-1}/(m-1)!$, for $m \geq 1$.

The main idea of the proposed methodology is to develop the model from a flexible nonparametric mixture structure for the collection of dose-dependent response distributions $f(R, \boldsymbol{y}^*, \boldsymbol{u}^* \mid m)$, and from that structure obtain dose-response inference for all endpoints of interest. To build the mixture model, consider first a generic dose level x. We then represent the multivariate response distribution through the following mixture

$$\int \text{Bin}\left(R \mid m, \pi(\gamma)\right) \prod_{k=1}^{m-R} \text{Bern}\left(y_k^* \mid \pi(\theta)\right) \text{N}\left(u_k^* \mid \mu, \varphi\right) \, \text{d}G_x(\gamma, \theta, \mu)$$

where $\pi(u) = \exp(u)/\{1 + \exp(u)\}$, $u \in \mathbb{R}$, denotes the logistic function, and G_x is the dose-dependent mixing distribution for the mixture kernel parameters (γ, θ, μ). The Binomial part of the kernel for $R \mid m$ accounts for the possibility of non-viable fetuses, whereas the product kernel part for $\boldsymbol{y}^*, \boldsymbol{u}^* \mid R, m$ accounts for the potential endpoints of the live pups. Although not explicitly built in the mixture kernel, dependence between the malformation responses, y_k^*, and weight responses, u_k^*, is induced by mixing on the parameters of their respective Bernoulli and normal kernel distributions. The mixing can be extended to the variance, φ, of the birth weight normal kernel. This approach sacrifices the ability to promote an increasing trend (in prior expectation) for one of the risk functions discussed in Section 2.2, and involves more complex MCMC model fitting. Therefore, to strike a balance between model flexibility and computational feasibility, we adopt the location normal mixture component for the continuous endpoint.

Next, we need a flexible prior for the mixing distribution G_x. Here, nonparametric countable mixing provides desirable flexibility over continuous mixtures, which are limited to symmetry and unimodality, and an appealing alternative to discrete finite mixtures, which typically require more complex methods for inference and prior specification. Discrete mixing is particularly important as it allows clustering that, in turn, yields more

precise inference than parametric hierarchical models; Fronczyk & Kottas (2014) includes an example where a discrete nonparametric Binomial mixture provides striking improvement in uncertainty quantification over a Beta-Binomial model.

The Dirichlet process (DP) is the most widely used nonparametric prior for discrete random mixing distributions. We will use DP(α, H_0) to denote the DP prior for random distribution H, defined in terms of a centering (base) distribution H_0, and precision parameter $\alpha > 0$. Using its constructive definition (Sethuraman, 1994), the DP prior generates countable mixtures of point masses with atoms drawn from the base distribution and weights defined by a stick-breaking process. Specifically, a random distribution, H, drawn from DP(α, H_0) has an almost sure representation as $H = \sum_{l=1}^{\infty} \omega_l \delta_{\eta_l}$, where δ_a denotes a point mass at a, the η_l are i.i.d. from H_0, and $\omega_1 = \zeta_1$, $\omega_l = \zeta_l \prod_{r=1}^{l-1}(1-\zeta_r)$, for $l \geq 2$, with ζ_l i.i.d. from a Beta(1,α) distribution (independently of the η_l).

Now, given a DP prior for the mixing distribution, G_x, we have a probabilistic model for the clustered discrete-continuous outcomes at a specific dose level, x. To complete the model specification for the collection of response distributions over the range of dose values, $\mathcal{X} \subset \mathbb{R}^+$, we seek a prior probability model for the collection of mixing distributions $G_{\mathcal{X}} = \{G_x : x \in \mathcal{X}\}$. The dependent Dirichlet process (DDP) prior (MacEachern, 2000) provides an attractive option for such modeling, since it yields general nonparametric dependence across dose levels while resulting in a DP prior for each G_x. Here, we utilize the "common-weights" DDP prior structure,

$$G_{\mathcal{X}} = \sum_{l=1}^{\infty} \omega_l \, \delta_{\eta_{l\mathcal{X}}}, \tag{1}$$

where the ω_l arise from the DP stick-breaking process and the $\eta_{l\mathcal{X}} = \{\eta_l(x) : x \in \mathcal{X}\}$ are independent realizations from a stochastic process $G_{0\mathcal{X}}$ over $\mathcal{X}$. Hence, the prior model

for $G_{\mathcal{X}}$ can be viewed as a countable mixture of realizations from the base stochastic process $G_{0\mathcal{X}}$, with weights matching those from a single DP. Applications of common-weights DDP mixture models include: ANOVA settings (DeIorio et al., 2004); spatial modeling (Gelfand et al., 2005); dynamic density estimation (Rodriguez & ter Horst, 2008); quantile regression (Kottas & Krnjajić, 2009); survival regression (DeIorio et al., 2009); extreme value analysis (Kottas et al., 2012); autoregressive time series modeling (DiLucca et al., 2013); and modeling for dynamic marked point process intensities (Xiao et al., 2015). Support properties of DDP prior models are studied in Barrientos et al. (2012).

We thus propose the following DDP mixture model for the collection of dose-dependent response distributions for clustered binary and continuous outcomes:

$$f(R, \boldsymbol{y}^*, \boldsymbol{u}^* \mid m, G_{\mathcal{X}}) = \int \text{Bin}\,(R \mid m, \pi(\gamma)) \prod_{k=1}^{m-R} \text{Bern}\,(y_k^* \mid \pi(\theta))\, \text{N}\,(u_k^* \mid \mu, \varphi)\; \text{d}G_{\mathcal{X}}(\gamma, \theta, \mu) \quad (2)$$

with $G_{\mathcal{X}} \mid \alpha, \boldsymbol{\psi} \sim \text{DDP}(\alpha, G_{0\mathcal{X}})$, extending the DP notation and letting $\boldsymbol{\psi}$ denote the parameters of the base stochastic process $G_{0\mathcal{X}}$. Note that the atoms of the DDP prior comprise three mixing components, i.e., $\eta_l(x) = (\gamma_l(x), \theta_l(x), \mu_l(x))$. We accordingly define $G_{0\mathcal{X}}$ through a product of three isotropic Gaussian processes (GPs) with linear mean functions. Specifically, the GP prior associated with γ has mean function, $\xi_0 + \xi_1 x$, variance τ^2, and correlation function $\exp\{-\rho|x - x'|\}$; the mean function for θ is $\beta_0 + \beta_1 x$, the variance σ^2, and the correlation function $\exp\{-\phi|x - x'|\}$; and the GP prior on μ includes mean function $\chi_0 + \chi_1 x$, variance ν^2, and correlation function $\exp\{-\kappa|x - x'|\}$. Thereby, the GP hyperparameters are given by $\boldsymbol{\psi} = (\xi_0, \xi_1, \tau^2, \rho, \beta_0, \beta_1, \sigma^2, \phi, \chi_0, \chi_1, \nu^2, \kappa)$.

As discussed in Section 2.2, the form of the GP mean functions is a key part of the model specification with respect to the implied dose-response curves. The choice of the exponential correlation functions is driven by simplicity taking into account the fact that the DDP prior generates non-stationary realizations (with non-Gaussian finite dimensional

distributions) even though it is centered around isotropic GPs. Smoothness properties of DDP realizations in the context of spatial modeling are discussed in Gelfand et al. (2005) and Guindani & Gelfand (2006). In particular, the continuity of the GP realizations that define $G_{0\mathcal{X}}$ yields that the difference between G_x and $G_{x'}$ gets *smaller* as the distance between dose levels x and x' gets smaller. In our context, the practical implication is that of a smooth evolution across dose values for the multivariate response distribution and for the dose-response relationships associated with the multiple endpoints of interest.

The full Bayesian model is completed with an inverse gamma prior for φ, a gamma hyperprior for α, and with (independent) hyperpriors for the components of $\boldsymbol{\psi}$. Specification of these hyperpriors is discussed in Section 3.1 in the context of the DYME data example. The technical details on the hierarchical model formulation for the data, the MCMC posterior simulation method, and the approach to predictive inference at dose levels outside the set of observed doses are found in the Supplementary Material.

2.2 Functionals for risk assessment

Of key importance is study of dose-response relationships for risk assessment. In addition to dose-response curves for prenatal death, malformation, and low birth weight, we obtain risk functions that combine different endpoints. Although the dose-response curves are not modeled directly, their form can be developed through the respective probabilities implied by the DDP mixture model for a generic implant (associated with a generic dam) at dose level x. (For simpler notation, the implicit conditioning on $m = 1$ is excluded from the expressions below.) Given an implant at dose x, the Binomial kernel component in (2) reduces to $\text{Bern}(R^* \mid \pi(\gamma))$ for the single prenatal death indicator, R^*, with the remainder of the mixture kernel, $\text{Bern}(y^* \mid \pi(\theta))\text{N}(u^* \mid \mu, \varphi)$, present when $R^* = 0$. More generally, model (2) can be equivalently expressed in terms of $(\boldsymbol{R}^*, \boldsymbol{y}^*, \boldsymbol{u}^*)$, where $\boldsymbol{R}^* =$

$\{R_s^* : s = 1, ..., m\}$ are binary prenatal death responses, by replacing the $\text{Bin}(R \mid m, \pi(\gamma))$ kernel with $\prod_{s=1}^{m} \text{Bern}(R_s^* \mid \pi(\gamma))$ and setting $R = \sum_{s=1}^{m} R_s^*$.

The first dose-response curve is for embryolethality, that is, the probability of a non-viable fetus across effective dose levels,

$$D(x) \equiv \Pr(R^* = 1 \mid G_x) = \int \pi(\gamma)\, \mathrm{d}G_x(\gamma), \quad x \in \mathcal{X}.$$

Provided $\xi_1 > 0$ in the linear mean function of the respective DDP centering GP, $D(x)$ is increasing in prior expectation. Specifically, $\mathrm{E}(D(x)) = \int \pi(\gamma) \mathrm{d}G_{0x}(\gamma)$, where $G_{0x}(\gamma) = \mathrm{N}(\gamma \mid \xi_0 + \xi_1 x, \tau^2)$. Since G_{0x} is stochastically ordered in x when $\xi_1 > 0$, and $\pi(\gamma)$ is an increasing function, $\mathrm{E}(D(x))$ is a non-decreasing function of x.

For the malformation endpoint, consider the conditional probability of the corresponding binary response given a viable fetus, $M(x) \equiv \Pr(y^* = 1 \mid R^* = 0, G_x) = \Pr(y^* = 1, R^* = 0 \mid G_x)/\Pr(R^* = 0 \mid G_x)$. Hence, the malformation dose-response curve is given by

$$M(x) = \frac{\int \{1 - \pi(\gamma)\} \pi(\theta)\, \mathrm{d}G_x(\gamma, \theta)}{\int \{1 - \pi(\gamma)\}\, \mathrm{d}G_x(\gamma)}, \quad x \in \mathcal{X}.$$

Regarding the continuous outcome, we consider two risk assessment functionals. The first involves the expected birth weight conditioning on a viable fetus, $\mathrm{E}(u^* \mid R^* = 0, G_x) = \int u^* f(R^* = 0, u^* \mid G_x) \mathrm{d}u^* / \{\Pr(R^* = 0 \mid G_x)\}$. Using the mixture representation for $f(R^* = 0, u^* \mid G_x)$, we obtain

$$\mathrm{E}(u^* \mid R^* = 0, G_x) = \frac{\int \{1 - \pi(\gamma)\} \mu\, \mathrm{d}G_x(\gamma, \mu)}{\int \{1 - \pi(\gamma)\}\, \mathrm{d}G_x(\gamma)}, \quad x \in \mathcal{X}.$$

Alternatively, we can quantify the risk of low birth weight through $\Pr(u^* < \mathcal{U} \mid R^* = 0, G_x)$, for any cutoff point, $\mathcal{U}$, that is deemed sufficiently small. Following the literature

(e.g., Regan & Catalano, 1999), we take the cutoff to be two standard deviations below the average birth weight at the control level. It can be shown that

$$\Pr(u^* < \mathcal{U} \mid R^* = 0, G_x) = \frac{\int \{1 - \pi(\gamma)\} \Phi((\mathcal{U} - \mu)/\varphi^{1/2}) \, \mathrm{d}G_x(\gamma, \mu)}{\int \{1 - \pi(\gamma)\} \, \mathrm{d}G_x(\gamma)}, \quad x \in \mathcal{X},$$

where $\Phi(\cdot)$ denotes the standard normal distribution function.

The combined risk of the discrete outcomes can be studied through the probability of embryolethality or malformation, $r_d(x) \equiv \Pr(R^* = 1 \text{ or } y^* = 1 \mid G_x) = \Pr(R^* = 0, y^* = 1 \mid G_x) + \Pr(R^* = 1 \mid G_x)$, which results in

$$r_d(x) = 1 - \int \{1 - \pi(\gamma)\}\{1 - \pi(\theta)\} \, \mathrm{d}G_x(\gamma, \theta), \quad x \in \mathcal{X}.$$

As with the embryolethality endpoint, it is possible to promote an increasing trend in $r_d(x)$ through $\mathrm{E}(r_d(x)) = 1 - \left[\int \{1 - \pi(\gamma)\} \mathrm{dN}(\gamma \mid \xi_0 + \xi_1 x, \tau^2)\right] \left[\int \{1 - \pi(\theta)\} \mathrm{dN}(\theta \mid \beta_0 + \beta_1 x, \sigma^2)\right]$, where we have used the assumption of independent GP components for $G_{0\mathcal{X}}$. Now, if $\xi_1 > 0$ and $\beta_1 > 0$, each of the integral terms above is non-increasing in x and thus $\mathrm{E}(r_d(x))$ is a non-decreasing function of x.

Finally, a full risk function can be built through the probability of either of the discrete endpoints or low birth weight. Specifically, $r_f(x) \equiv \Pr(R^* = 1 \text{ or } y^* = 1 \text{ or } u^* < \mathcal{U} \mid G_x) = r_d(x) + \Pr(R^* = 0, y^* = 0, u^* < \mathcal{U} \mid G_x)$, which thus separates the effect of the negative outcomes from the discrete and continuous endpoints. Using the expression for $r_d(x)$ and the mixture form for $f(R^* = 0, y^* = 0, u^* \mid G_x)$, we can write

$$r_f(x) = 1 - \int \{1 - \pi(\gamma)\}\{1 - \pi(\theta)\}\{1 - \Phi((\mathcal{U} - \mu)/\varphi^{1/2})\} \, \mathrm{d}G_x(\gamma, \theta, \mu), \quad x \in \mathcal{X}.$$

Extending the argument above for $\mathrm{E}(r_d(x))$, it can be shown that $\mathrm{E}(r_f(x))$ is also a non-

decreasing function of x provided $\xi_1 > 0$, $\beta_1 > 0$, and $\chi_1 < 0$.

The above restriction on the slope parameters for the GP linear mean functions is readily implemented through their hyperpriors. We can thus promote increasing trends, through their prior expectation, in the embryolethality dose-response curve and in both combined risk functions. Although the argument does not extend to the conditional probability of malformation or low birth weight, extensive prior simulations suggest that the $\xi_1 > 0$, $\beta_1 > 0$, and $\chi_1 < 0$ restrictions induce non-decreasing prior expectations also for these dose-response curves. Given the small number of observed doses in developmental toxicology data, this level of structure in the prior is key for practicable inference for the multiple dose-response relationships, since such inference requires interpolation and extrapolation beyond the observed dose levels. However, the modeling approach does not imply (with prior probability 1) monotonic dose-response relationships. This is a practically important feature for toxicity experiments that may depict a hormetic effect. Hormesis refers to a dose-response phenomenon characterized by favorable biological responses to low exposures to toxins (e.g., Calabrese, 2005). For endpoints involving mutation, birth defects, or cancer, hormesis may result in non-monotonic, J-shaped dose-response curves. Fronczyk & Kottas (2014) study an example that involves a non-monotonic dose-response relationship, under the simpler data setting without continuous outcomes.

As a further inferential goal, we investigate different types of intra-litter correlations, i.e., correlations for two live pups within the same litter at dose x. In particular, we obtain inference for the correlation between: the discrete malformation endpoints, $\text{Corr}(y_k^*, y_{k'}^* \mid R_k^* = 0, R_{k'}^* = 0, G_x)$; the continuous endpoints, $\text{Corr}(u_k^*, u_{k'}^* \mid R_k^* = 0, R_{k'}^* = 0, G_x)$; and the weight and malformation endpoints, $\text{Corr}(y_k^*, u_{k'}^* \mid R_k^* = 0, R_{k'}^* = 0, G_x)$. Of interest is also the intra-fetus correlation between the discrete and continuous outcomes for one viable fetus, $\text{Corr}(y^*, u^* \mid R^* = 0, G_x)$. (The above expressions involve conditioning

on either $m = 2$ or $m = 1$, which is again suppressed from the notation.) Although parametric hierarchical models can be extended to accommodate dose-dependent intra-litter correlations, the regression formulation for the dependence on dose is not trivial to specify. Through flexible modeling for the multivariate response distributions, the DDP mixture in (2) yields dose-dependent nonparametric inference for the association among the clustered discrete-continuous outcomes.

Some of the expectations required for the correlation expressions have been obtained, e.g., $M(x) = \mathrm{E}(y^* \mid R^* = 0, G_x) = \mathrm{E}(y^{*2} \mid R^* = 0, G_x)$. The remaining expectations can be developed using similar derivations. For instance, $\mathrm{E}(y^*_k u^*_{k'} \mid R^*_k = 0, R^*_{k'} = 0, G_x) = \int\{1-\pi(\gamma)\}^2\pi(\theta)\mu\,\mathrm{d}G_x(\gamma,\theta,\mu)/[\int\{1-\pi(\gamma)\}^2\mathrm{d}G_x(\gamma)]$, and $\mathrm{E}(y^*u^* \mid R^* = 0, G_x)$ is given by an analogous expression substituting $\{1-\pi(\gamma)\}^2$ by $\{1-\pi(\gamma)\}$.

3 Data illustration

We use the data discussed in the Introduction to demonstrate the practical utility of the model. Conducted by the National Toxicology Program, the particular toxicity study investigates the organic solvent diethylene glycol dimethyl ether (DYME). There are five observed dose levels, one control and four active (62.5, 125, 250, and 500 mg/kg). The number of animals exposed to each level ranges from 18 to 24, and the number of implants across all doses ranges from 3 to 17, with 25th, 50th, and 75th percentiles of 12, 13, and 14, respectively. More details on the data can be found in the Supplementary Material.

3.1 Prior specification

The DDP mixture model is implemented with an inverse gamma prior for φ with shape parameter 2 and mean 1, a gamma$(2, 1)$ prior for α, and independent hyperpriors assigned

to the parameters of the centering GPs. More specifically, we use uniform priors on $(0, B)$ for the GP correlation parameters ρ, ϕ and κ; inverse gamma priors for the GP variances τ^2, σ^2 and ν^2 (with shape parameters equal to 2, implying infinite prior variance); and normal priors for the intercepts of the linear mean functions ξ_0, β_0 and χ_0. Moreover, to incorporate the prior structure for the dose-response curves discussed in Section 2.2, we place exponential priors on ξ_1 and β_1, and a normal prior on χ_1 truncated above at 0.

We specify B using the *range of dependence* interpretation for the GP exponential correlation function. For instance, for the first GP component of $G_{0\mathcal{X}}$, $3/\rho$ is the distance between dose levels that yields correlation 0.05. The range of dependence is usually assumed to be a fraction of the maximum interpoint distance over the index space. Hence, since $3/B < 3/\rho$, we specify B such that $3/B = rd_{\max}$, for small r, where $d_{\max}$ is the maximum distance between observed doses; $B = 1$ was used for the DYME data analysis. The remaining hyperprior parameters are chosen to provide dispersed prior distributions for the implied dose-response relationships. In particular, for the DYME data, the prior mean for $D(x)$ and for $M(x)$ begins around 0.5 and has a slight increasing trend, and the corresponding 95% prior uncertainty bands essentially span the $(0, 1)$ interval. In addition, the prior distribution for $\mathrm{E}(u^* \mid R^* = 0, G_x)$ across dose levels is centered around 1 g and spans from about 0 g to 2 g; note that healthy pups weigh $0.5 - 1.5$ g. A plausible range, R_w, of birth weight values can also be used to set the prior mean for φ through, for instance, $(R_w/4)^2$. Finally, parameter α controls the number of distinct components in the DP mixture model for the data induced by the common-weights DDP prior, which can be used to guide the choice of the gamma prior for α.

Although this approach does not uniquely specify all the hyperpriors, it offers a practical strategy to complete the DDP mixture model specification based on a small amount of prior information. Note that all that is required is a rough guess at the number of distinct

mixture components, a range for the dose values of interest, and a range of values for the continuous outcome at the control. Interestingly, despite the moderate sample sizes of the DYME data, there is substantial learning for all DDP prior hyperparameters with posterior densities significantly concentrated relative to the corresponding prior densities.

3.2 Risk assessment inference results

Figure 2 plots the posterior mean and 90% uncertainty bands of the dose-response curves. The probability of embryolethality depicts an increasing trend. The conditional probability of malformation has a skewed shape, with larger uncertainty between the last two observed dose levels, 250 and 500 mg/kg. The combined risk function for the discrete outcomes is similar in shape to the malformation dose-response curve, though shifted up slightly and with decreased uncertainty bands. The expected birth weight curve has a relatively constant decreasing rate. The probability of low fetal weight (where the cutoff is 0.782 g) reveals an increasing exponential trend with wider uncertainty bands as dose level increases. Finally, the full risk function is not substantially shifted up relative to the combined risk of the discrete outcomes, suggesting that the embryolethality and malformation endpoints are the main contributors to the overall dose-response relationship.

Figure 3 shows inference results at the active dose levels for the different types of correlations discussed in Section 2.2. The posterior densities for the intra-fetus correlation between the malformation and weight outcomes are concentrated around zero, other than for level 250 mg/kg where a mild negative correlation is suggested. Although not shown here, the results were similar for the intra-litter correlation between the malformation and weight endpoints. There is little intra-litter correlation between the malformation responses at the two smaller active dose levels, whereas increasing the level at 250 mg/kg increases the correlation such that if one pup exhibits a malformation it is likely another

pup within that litter will also show birth defects. At dose 500 mg/kg, the rate of embryolethality is largest and, thus, not many implants grow enough to develop birth defects. This limited amount of information from the data may explain the dispersed density for the corresponding correlation. Finally, the posterior densities for the intra-litter correlation between the fetal weight outcomes are roughly similar at the four active toxin levels and concentrated mainly on positive values, which reflects that birth weight of a generic pup affects another pup within the litter in a similar fashion.

Also of interest is inference for the various response distributions. In Figure 4, we report posterior mean estimates for: probability mass functions of the number of non-viable fetuses given a specific number of implants; probability mass functions of the number of malformations given a specified number of implants and non-viable fetuses; and fetal weight densities. More comprehensive results, including point and interval estimates at all observed dose levels and at new doses, are provided in the Supplementary Material.

3.3 Model assessment

As an approach to model checking, we examine cross-validation posterior predictive residuals. For any observed dose level x_0, the joint posterior predictive distribution for new number of implants, m_0, number of prenatal deaths, R_0, malformation responses, $\boldsymbol{y}_0^* = \{y_{0k}^* : k = 1, \ldots, m_0 - R_0\}$, and fetal weight outcomes, $\boldsymbol{u}_0^* = \{u_{0k}^* : k = 1, \ldots, m_0 - R_0\}$, can be factorized into $p(m_0 \mid \text{data}) = \int f(m_0 \mid \lambda) \mathrm{d}p(\lambda \mid \text{data})$, and $p(R_0, \boldsymbol{y}_0^*, \boldsymbol{u}_0^* \mid m_0, \text{data}) = \int f(R_0, \boldsymbol{y}_0^*, \boldsymbol{u}_0^* \mid m_0, G_{x_0}) \mathrm{d}p(\Theta \mid \text{data})$. Here, Θ represents all parameters of the DDP mixture model $f(R_0, \boldsymbol{y}_0^*, \boldsymbol{u}_0^* \mid m_0, G_{x_0})$, which arises from (2) applied at the specific dose x_0. The posterior predictive distribution can be extended to the entire vector of observed dose levels (as well as to new dose values).

We use one, randomly chosen, cross-validation sample comprising data from 20 dams

(approximately 20% of the data) spread roughly evenly across the dose levels. After fitting the DDP mixture model to the reduced DYME data, we obtain for each observed toxin level posterior predictive samples for R_0/m_0, $(m_0 - R_0)^{-1} \sum_{k=1}^{m_0-R_0} y_{0k}^*$, and $(m_0 - R_0)^{-1} \sum_{k=1}^{m_0-R_0} u_{0k}^*$. Figure 5 includes box plots of these samples along with the corresponding values from the cross-validation data points. This graphical model checking approach gives no strong evidence of ill-fitting.

3.4 Comparison with a parametric hierarchical model

To demonstrate the benefits of the nonparametric mixture model relative to simpler model specifications, we implement a hierarchical model built from parametric distributions and dose-response curves. We consider the commonly utilized setting for pup-specific outcomes modelled through a bivariate normal distribution for the weight outcomes and for latent continuous malformation responses (e.g., Regan & Catalano, 1999; Faes et al., 2006).

Following the notation of Section 2.1, let $\boldsymbol{y}^{**} = \{y_k^{**} : k = 1, \ldots, m - R\}$ be latent continuous malformation responses, such that $y_k^* = 1$ if and only if $y_k^{**} > 0$. Using again a shifted Poisson distribution for m, the parametric response distribution is defined by

$$f(R, \boldsymbol{y}^{**}, \boldsymbol{u}^* \mid m) = \int \mathrm{Bin}(R \mid m, \pi) \prod_{k=1}^{m-R} \mathrm{N}_2((y_k^{**}, u_k^*) \mid \lambda, \mu, \rho, \sigma_u^2) \mathrm{d}H_x(\pi, \lambda, \mu)$$

where $\mathrm{N}_2(\cdot \mid \lambda, \mu, \rho, \sigma_u^2)$ is a bivariate normal distribution with mean vector (λ, μ), correlation parameter ρ, variance corresponding to u_k^* given by σ_u^2, and variance corresponding to y_k^{**} fixed at 1 for identifiability. The parametric mixing (random-effects) distribution H_x comprises independent components: $\mathrm{Beta}(\zeta\pi(\beta_0 + \beta_1 x), \zeta\{1 - \pi(\beta_0 + \beta_1 x)\})$ for π, $\mathrm{N}(\lambda_0 + \lambda_1 x, \sigma_\lambda^2)$ for λ, and $\mathrm{N}(\mu_0 + \mu_1 x, \sigma_\mu^2)$ for μ, such that the parameter vector includes ρ, σ_u^2, and the random-effects distribution hyperparameters, $(\zeta, \beta_0, \beta_1, \lambda_0, \lambda_1, \sigma_\lambda^2, \mu_0, \mu_1, \sigma_\mu^2)$.

Regarding risk assessment functionals, we consider as in Section 2.2 a generic implant at dose level x. The prenatal death indicator is R^*, and when $R^* = 0$, we observe the binary malformation outcome y^*, and the fetal weight response u^*. If in the model for (R^*, y^{**}, u^*) we first integrate out (π, λ, μ) and then y^{**}, the model for (R^*, y^*, u^*) becomes $\text{Bern}(R^* \mid \pi(\beta_0+\beta_1 x))\text{N}(u^* \mid \mu_0+\mu_1 x, \sigma_u^2+\sigma_\mu^2)\text{Bern}(y^* \mid q(x,u^*))$, where $q(x,u^*) = \Phi\left((1-\rho^{*2})^{-1/2}\{(1+\sigma_\lambda^2)^{-1/2}(\lambda_0+\lambda_1 x) + \rho^*(\sigma_u^2+\sigma_\mu^2)^{-1/2}(u^*-\mu_0-\mu_1 x)\}\right)$. Here, $\rho^* = \rho\sigma_u(1+\sigma_\lambda^2)^{-1/2}(\sigma_u^2+\sigma_\mu^2)^{-1/2}$ is the intra-fetus correlation between the latent malformation response and the weight outcome. Based on this distribution for (R^*, y^*, u^*), the embryolethality and malformation dose-response curves are given by $\Pr(R^*=1) = \pi(\beta_0+\beta_1 x)$ and $\Pr(y^*=1 \mid R^*=0) = \int q(x,u^*)\text{N}(u^* \mid \mu_0+\mu_1 x, \sigma_u^2+\sigma_\mu^2)\,\text{d}u^*$. For the fetal weight endpoint, we obtain $\text{E}(u^* \mid R^*=0) = \mu_0+\mu_1 x$, and $\Pr(u^* < \mathcal{U} \mid R^*=0) = \Phi(\{\mathcal{U}-\mu_0-\mu_1 x\}/\{\sigma_u^2+\sigma_\mu^2\}^{1/2})$. Intra-litter correlations are derived from the model distribution for two live pups within the same litter at dose x. The correlation among latent malformation responses is given by $\text{Corr}(y_k^{**}, y_{k'}^{**} \mid R_k^*=0, R_{k'}^*=0) = \sigma_\lambda^2/(1+\sigma_\lambda^2)$, and the correlation among weight outcomes by $\text{Corr}(u_k^*, u_{k'}^* \mid R_k^*=0, R_{k'}^*=0) = \sigma_\mu^2/(\sigma_u^2+\sigma_\mu^2)$. Note that $\text{Corr}(y_k^{**}, u_{k'}^* \mid R_k^*=0, R_{k'}^*=0) = 0$, although this correlation will be non-zero under a dependent random-effects distribution for (λ, μ).

We implement the parametric hierarchical model for the DYME data, using an MCMC algorithm which imputes the latent malformation responses based on their truncated normal full conditional distributions. Given the latent responses, standard Gibbs sampling updates are available for the random effects parameters, as well as for λ_0, λ_1, σ_λ^2, μ_0, μ_1, and σ_μ^2. The remaining hyperparameters, ζ, β_0, and β_1, as well as ρ and σ_u^2 are sampled with Metropolis-Hastings steps. Priors for model parameters were specified such that prior point and interval estimates for the various dose-response curves were comparable with the corresponding prior estimates under the DDP mixture model.

Figure 6 contrasts inference results for three dose-response curves under the parametric and nonparametric models. Evidently, the DDP mixture model is more successful in uncovering the dose-response relationships suggested by the data, the most striking difference with the parametric model arising for the probability of low birth weight.

We also consider more formal model comparison based on the posterior predictive loss criterion of Gelfand & Ghosh (1998), applied to each of the endpoints in the same spirit with Section 3.3. Let $j = 1, ..., n_i$ index the dams at observed dose x_i, for $i = 1, ..., N$. For each x_i, we draw replicate responses $\tilde{m}_i$, $\tilde{R}_i$, $\tilde{\boldsymbol{y}}_i^* = \{\tilde{y}_{ik}^* : k = 1, \ldots, \tilde{m}_i - \tilde{R}_i\}$, and $\tilde{\boldsymbol{u}}_i^* = \{\tilde{u}_{ik}^* : k = 1, \ldots, \tilde{m}_i - \tilde{R}_i\}$. (Note that the responses from the n_i dams at the ith dose level share the same *covariate*, x_i, and we thus need one posterior predictive sample at each dose.) For the DDP mixture model, these posterior predictive samples are obtained as discussed in Section 3.3. Sampling from the required posterior predictive distribution of the parametric model is also straightforward given its hierarchical structure. Then, for the embryolethality endpoint, the criterion favors the model $\mathcal{M}$ that minimizes the (possibly weighted) sum of $P(\mathcal{M}) = \sum_{i=1}^{N} n_i \mathrm{Var}(\tilde{R}_i/\tilde{m}_i \mid \text{data})$, a penalty term for model complexity, and $G(\mathcal{M}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \{R_{ij}/m_{ij} - \mathrm{E}(\tilde{R}_i/\tilde{m}_i \mid \text{data})\}^2$, a goodness-of-fit term. Here, m_{ij} and R_{ij} are the data values for the number of implants and the number of prenatal deaths, respectively, from the jth dam at dose x_i. The $P(\mathcal{M})$ and $G(\mathcal{M})$ components are defined analogously for the malformation endpoint, based on predictive samples $(\tilde{m}_i - \tilde{R}_i)^{-1} \sum_{k=1}^{\tilde{m}_i - \tilde{R}_i} \tilde{y}_{ik}^*$, and for the average birth weight endpoint, using predictive samples $(\tilde{m}_i - \tilde{R}_i)^{-1} \sum_{k=1}^{\tilde{m}_i - \tilde{R}_i} \tilde{u}_{ik}^*$. The results, reported in Table 1, favor the nonparametric model across all endpoints. For the embryolethality endpoint, the DDP mixture model fares better with respect to both criterion components. In the other two cases, the penalty term is slightly smaller for the parametric model, but the DDP mixture model results in a substantially smaller goodness-of-fit term.

The DYME data analysis demonstrates the benefits of the nonparametric model formulation, and the challenges for parametric modeling in this application area which requires specification of a multivariate response distribution for mixed clustered outcomes, as well as of various dose-response functions. It is of course possible to increase the flexibility of the parametric model considered here, by extending the random-effects distribution for (λ, μ) to include additional polynomial terms in the means, $\lambda_0 + \lambda_1 x$ and $\mu_0 + \mu_1 x$, and/or to enable dose-dependent intra-litter correlations through dose-dependent variances, σ^2_λ and σ^2_μ. However, more general parametric dose-response functions become increasingly more difficult to select, especially for the variance components, and they can substantially complicate MCMC model fitting. In this respect, the proposed DDP mixture model is arguably attractive as it can be implemented with a posterior simulation algorithm which is not more complicated than the ones for general parametric models, while at the same time allowing the flexibility in distributional and dose-response function shapes provided by the large support of the nonparametric prior.

4 Discussion

The approach developed here is applicable to developmental toxicity experiments involving clustered categorical outcomes and continuous responses. The modeling framework provides flexibility in the multiple response distributions as well as the various risk assessment quantities. The proposed model involves DDP mixing with respect to three parameters. This results in a relatively complex setting for prior specification and posterior simulation. However, the data analysis results demonstrate that the methodology is feasible to implement given sufficient amounts of data, as well as that it can lead to substantial improvements in predictive inference relative to parametric hierarchical models.

It is worth noting that the common-weights DDP prior structure is particularly well suited for nonparametric mixture modeling in the context of developmental toxicity studies. The general DDP version, $G_{\mathcal{X}} = \sum_{l=1}^{\infty} \omega_{l\mathcal{X}} \delta_{\eta_{l\mathcal{X}}}$, with both weights and atoms evolving across dose level, is impractical for this application area, since the typical developmental toxicity experiment involves collections of responses at a small number of dose levels. Contrarily to the common-weights DDP simplification, we may have chosen a prior structure where only the weights evolve with dose, that is, $G_{\mathcal{X}} = \sum_{l=1}^{\infty} \omega_{l\mathcal{X}} \delta_{\eta_l}$. Here, the $\eta_l = (\gamma_l, \theta_l, \mu_l)$ are i.i.d. from a base distribution G_0, independently of the stochastic mechanism that generates the $\omega_{l\mathcal{X}} = \{\omega_l(x) : x \in \mathcal{X}\}$. Given the relatively large number of mixing parameters in model (2), this prior structure appears on the surface to be a more suitable simplification of the general DDP prior. However, this "common-atoms" DDP formulation presents a formidable complication with regard to anchoring the inference for the various dose-response relationships through a monotonic trend in prior expectation (see Section 2.2). For instance, under a common-atoms DDP prior, it can be shown for the embryolethality dose-response curve that $\mathrm{E}(D(x)) = \int \pi(\gamma) \mathrm{d}G_0(\gamma)$, which is constant in x rendering interpolation and extrapolation inference practically useless.

Finally, it may be useful to entertain simpler versions of model (2). A possible semiparametric version excludes the distribution of prenatal deaths from the DDP mixing, building the response distribution through $f(m) f_x(R \mid m) f(\boldsymbol{y}^*, \boldsymbol{u}^* \mid R, m)$, where $f_x(R \mid m)$ is a parametric distribution, such as a Beta-Binomial with a logistic form for the probability of embryolethality. Now, the DDP mixture would be reserved for the pup-specific responses, $f(\boldsymbol{y}^*, \boldsymbol{u}^* \mid R, m, G_{\mathcal{X}}) = \int \prod_{k=1}^{m-R} \mathrm{Bern}(y_k^* \mid \pi(\theta)) \mathrm{N}(u_k^* \mid \mu, \varphi) \mathrm{d}G_{\mathcal{X}}(\theta, \mu)$. This model results in a simplified MCMC algorithm. Also, the conditional probabilities of malformation and low birth weight can be specified to be increasing in prior expectation. The downside is that the parametric form for the prenatal death distribution may not be

sufficiently flexible, as demonstrated in Section 3.4 with the DYME data.

Acknowledgments

The work of the second author was supported in part by the National Science Foundation under award DMS 1310438. The authors wish to thank an Associate Editor and two reviewers for useful feedback and for comments that improved the presentation of the material in the paper.

References

Barrientos, A. F., Jara, A. & Quintana, F. A. (2012). On the support of MacEachern's dependent Dirichlet processes and extensions. *Bayesian Analysis* **7**, 277–310.

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.

Catalano, P. & Ryan, L. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* **87**, 651–658.

DeIorio, M., Johnson, W. O., Müller, P. & Rosner, G. L. (2009). Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics* **63**, 762–771.

DeIorio, M., Müller, P., Rosner, G. & MacEachern, S. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association* **99**, 205–215.

DiLucca, M., Guglielmi, A., Müller, P. & Quintana, F. (2013). A simple class of Bayesian nonparametric autoregressive models. *Bayesian Analysis* **8**, 63–88.

Dominici, F. & Parmigiani, G. (2001). Bayesian semiparametric analysis of developmental toxicology data. *Biometrics* **57**, 150–157.

Dunson, D., Chen, Z. & Harry, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* **59**, 521–530.

Faes, C., Geys, H., Aerts, M. & Molenberghs, G. (2006). A hierarchical modeling approach for risk assessment in developmental toxicity studies. *Computational Statistics & Data Analysis* **51**, 1848–1861.

Fronczyk, K. & Kottas, A. (2014). A Bayesian nonparametric modeling framework for developmental toxicity studies (with discussion). *Journal of the American Statistical Association* **109**, 873–893.

Gelfand, A. & Ghosh, S. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85**, 1–11.

Gelfand, A., Kottas, A. & MacEachern, S. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.

Gueorguieva, R. & Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* **96**, 1102–1112.

Guindani, M. & Gelfand, A. E. (2006). Smoothness properties and gradient analysis under spatial Dirichlet process models. *Methodology and Computing in Applied Probability* **8**, 159–189.

Hwang, B. S. & Pennell, M. L. (2013). Semiparametric Bayesian joint modeling of a binary and continuous outcome with applications in toxicological risk assessment. *Statistics in Medicine* **33**, 1162–1175.

Kottas, A. & Fronczyk, K. (2013). Flexible Bayesian modelling for clustered categorical responses in developmental toxicology. In *Bayesian Theory and Applications*, Eds. P. Damien, P. Dellaportas, N. G. Polson & D. A. Stephens, pp. 70–83. Oxford University Press.

Kottas, A. & Krnjajić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics* **36**, 297–319.

Kottas, A., Wang, Z. & Rodríguez, A. (2012). Spatial modeling for risk assessment of extreme values from environmental time series: A Bayesian nonparametric approach. *Environmetrics* **23**, 649–662.

MacEachern, S. (2000). Dependent Dirichlet processes. Technical report, Ohio State University, Deptartment of Statistics.

Nott, D. & Kuk, A. (2009). Analysis of clustered binary data with unequal cluster sizes: A semiparametric Bayesian approach. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 101–118.

Price, C., Kimmel, C., George, J. & Marr, M. (1987). The developmental toxicity of diethylene glycol dimethyl ether in mice. *Fundamentals of Applied Pharmacology* **8**, 115–126.

Regan, M. & Catalano, P. (1999). Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. *Biometrics* **55**, 760–768.

Rodriguez, A. & ter Horst, E. (2008). Bayesian dynamic density estimation. *Bayesian Analysis* **3**, 339–366.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Xiao, S., Kottas, A. & Sansó, B. (2015). Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences. *The Annals of Applied Statistics* **9**, 353–382.

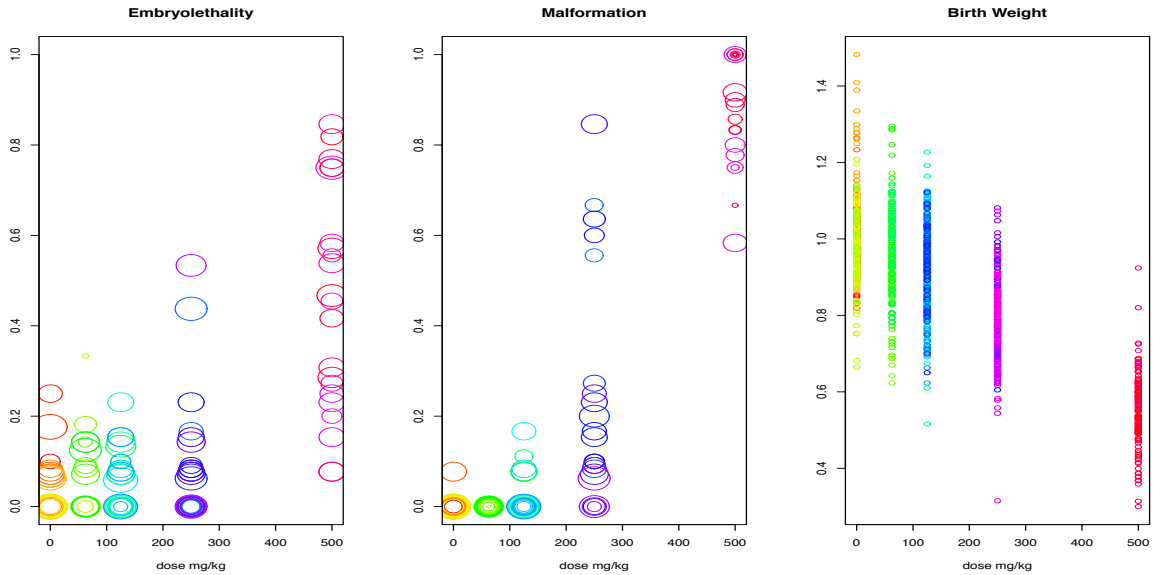

Figure 1: For the DYME data, the proportion of non-viable fetuses among implants for each dam at each dose level (left panel), the proportion of malformed pups among the viable fetuses for each dam at each dose level (middle panel), and the birth weights (in grams) of the live pups at each dose level (right panel). In the left and middle panels, each circle corresponds to a particular dam and the size of the circle is proportional to the number of implants and number of viable fetuses, respectively.

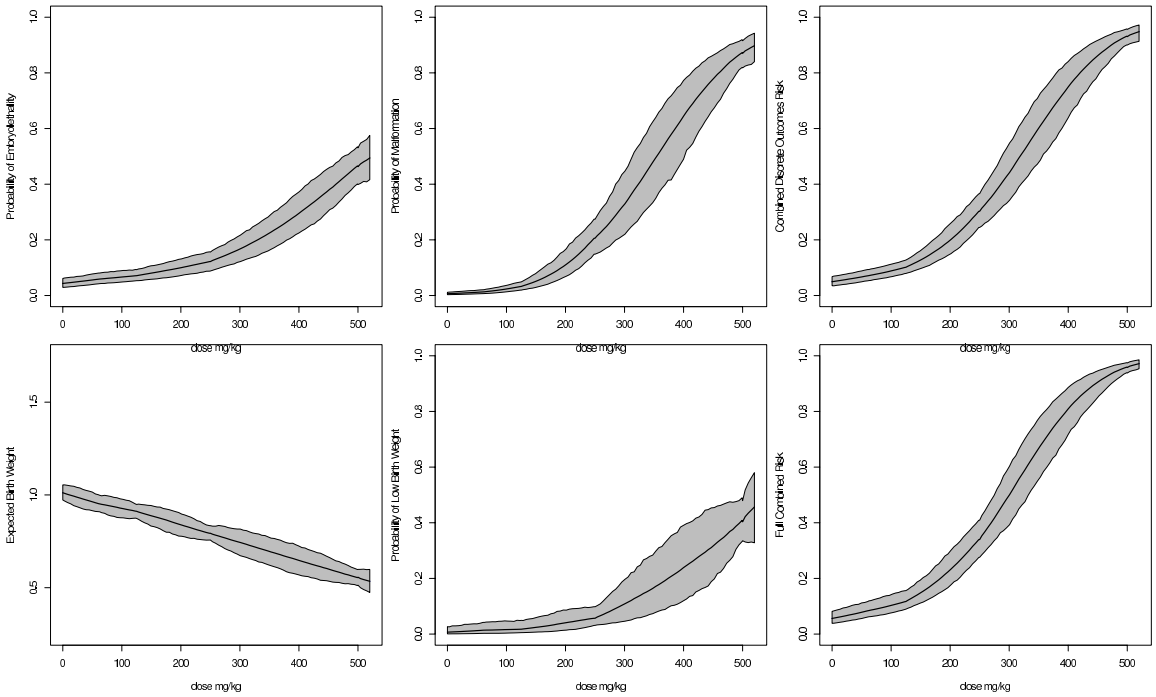

Figure 2: Posterior mean (solid line) and 90% uncertainty bands (gray shaded region) for: the probability of a non-viable fetus (top left); the conditional probability of malformation (top middle); the risk of combined discrete endpoints (top right); the expected birth weight (bottom left); the conditional probability of low birth weight (bottom middle); and the full combined risk function (bottom right).

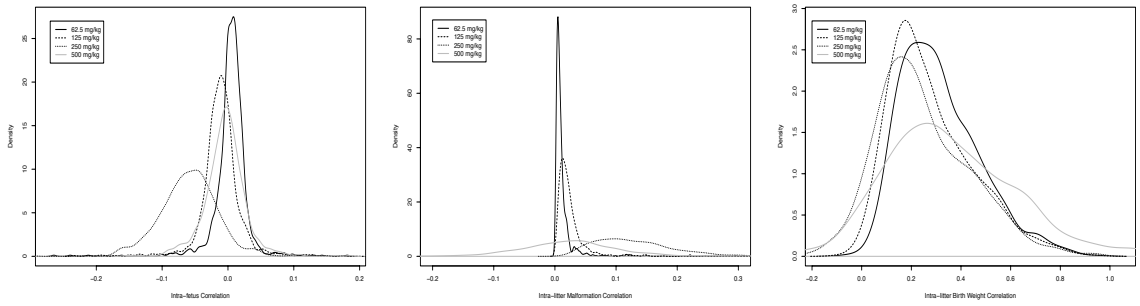

Figure 3: Posterior densities of the intra-fetus correlation between the malformation and weight outcomes (left panel), of the intra-litter correlation between the malformation responses (middle panel), and of the intra-litter correlation between the weight outcomes (right panel). In each case, results are shown for the four active toxin levels.

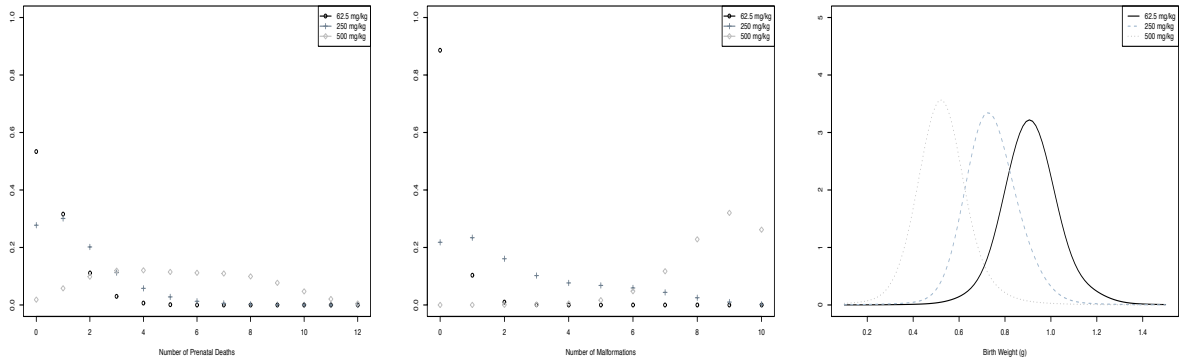

Figure 4: Posterior mean estimates for: the probability mass function of the number of non-viable fetuses given $m = 12$ implants, $f(R \mid m = 12, G_x)$ (left panel); the probability mass function of the number of malformations given $m = 12$ implants and $R = 2$ non-viable fetuses, $f(\sum_{k=1}^{m-R} y_k^* \mid m = 12, R = 2, G_x)$ (middle panel); and the probability density function for fetal weight, $f(u^* \mid m = 1, R = 0, G_x)$ (right panel). In each case, results are shown for three observed dose levels.

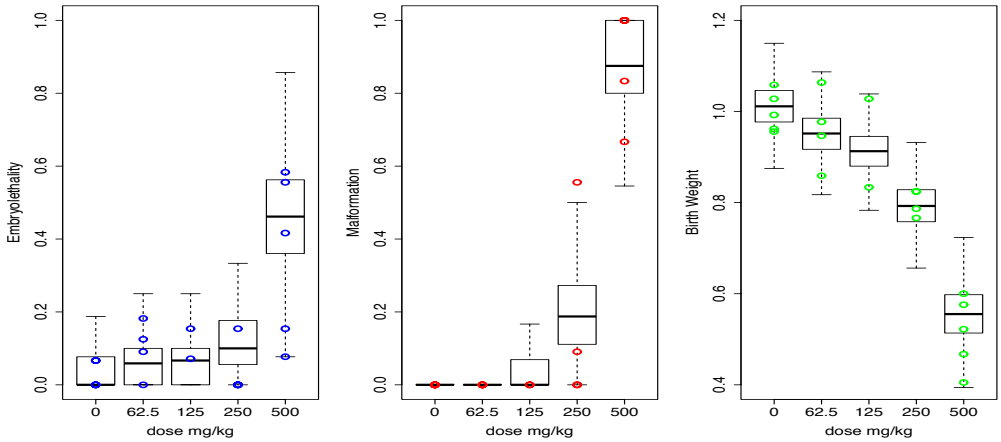

Figure 5: Box plots of posterior predictive samples for R_0/m_0 (left panel), $(m_0 - R_0)^{-1} \sum_{k=1}^{m_0-R_0} y_{0k}^*$ (middle panel), and $(m_0 - R_0)^{-1} \sum_{k=1}^{m_0-R_0} u_{0k}^*$ (right panel) at the five observed dose levels. The corresponding values from the 20 cross-validation data points are denoted by "o".

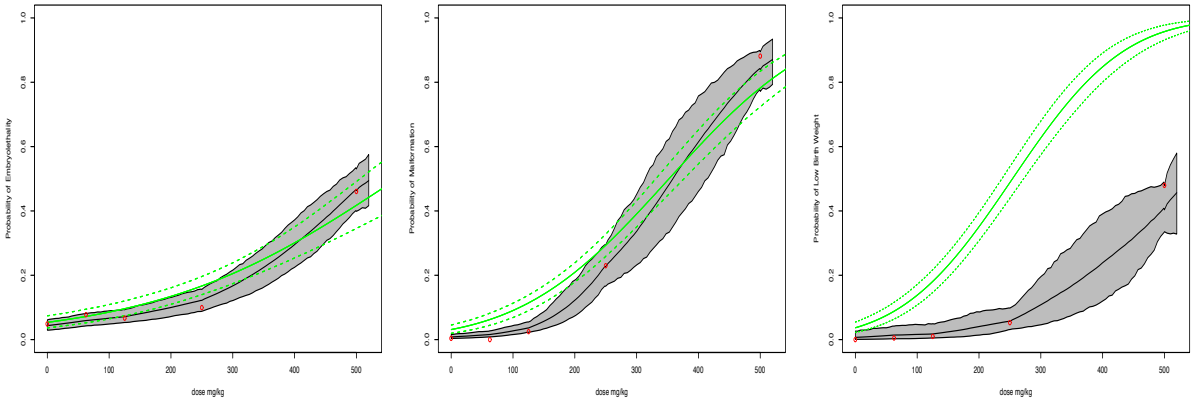

Figure 6: Comparison between the parametric and DDP mixture models. Posterior mean and 90% uncertainty bands for: the probability of a non-viable fetus (left panel); the conditional probability of malformation (middle panel); and the conditional probability of low birth weight (right panel). The interval estimates for the parametric model and the DDP mixture model are denoted by the dashed lines and the gray shaded region, respectively. Each panel includes the corresponding data-based estimates (denoted by "o") at the observed dose levels.

Endpoint	Goodness-of-fit term	Penalty term
Embryolethality	**2.07** 2.46	**1.04** 4.88
Malformation	**1.77** 12.7	**0.895** 0.778
Average birth Weight	**0.847** 45.4	**0.167** 0.121

Table 1: Comparison between the parametric and DDP mixture models. Values for the goodness-of-fit and penalty terms of the posterior predictive loss criterion for each of the embryolethality, malformation, and average birth weight endpoints. The values under the nonparametric DDP mixture model are given in bold.

Risk assessment for toxicity experiments with discrete and continuous outcomes: A Bayesian nonparametric approach

Supplementary material

Kassandra Fronczyk and Athanasios Kottas *

DYME data set

Developmental toxicity studies are designed to assess potential adverse effects of drugs and other exposures on developing fetuses. A typical experiment consists of four or five dose groups of 20 to 30 pregnant females, one group serving as a control and the others exposed to increasing levels of the toxin. Outcomes include fetal deaths and resorptions, several malformation indicators, and weight reduction. For illustrative purposes, we focus on one of many studies available from the National Toxicology Program database; the standard clustered multivariate response example found in the literature is one where diethylene

*K. Fronczyk is in the Applied Statistics Group at Lawrence Livermore National Laboratory, Livermore, CA, USA. (E-mail: *fornczyk1@llnl.gov*), and A. Kottas is Professor of Statistics, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, USA. (E-mail: *thanos@soe.ucsc.edu*). This research is part of the Ph.D. dissertation of the first author completed at University of California, Santa Cruz. The work of the second author was supported in part by the National Science Foundation under award DMS 1310438.

glycol dimethyl ether (DYME), an organic solvent, is evaluated for toxic effects in pregnant mice (Price et al., 1987). This example (see Table 1) shows clear dose-related reductions in fetal weight with the highest concentration of DYME resulting in roughly one-half the mean weight in control animals. The malformation data also suggest trends with dose; variations exhibit increases at thelower doses, giving way to strong dose-related trends in full malformations at the highest doses. The combined variation plus malformation outcome increases monotonically with dose level.

Table 1: Numerical summaries for the DYME data.

Dose	Dam	Implant	Fetal	Malformations	Litter size		Birth Weight	
(mg/kg)	count	count	deaths		Mean	SD	Mean	SD
0	21	296	14	1	13.4	2.4	1.00	0.11
62.5	18	214	14	0	11.1	3.4	0.96	0.12
125	24	312	21	7	12.1	2.3	0.91	0.11
250	23	299	32	59	11.6	2.1	0.79	0.10
500	22	275	127	128	6.7	3.2	0.55	0.10

Discussion of alternative prior probability models

The structure of the model as a nonparametric *discrete* mixture is essential. Such a structure utilizes the flexibility of mixture modeling, while at the same time, the discrete nature of the DDP-induced mixing distributions yields more effective modeling than parametric continuous mixtures. Nevertheless, it is of interest to entertain alternative nonparametric models directly for the dose-dependent response distributions. For example, a dependent Pólya tree prior could be such an alternative, although it is not clear how far this approach can be taken with respect to the extension to joint modeling for different outcomes. Although not considered other priors here, these priors generally do not offer significant practical advantages relative to the DP when the focus is on flexible inference for random distributions (as in our work) rather than on clustering inference in practice.

Another prior found in the literature is the kernel stick-breaking process Dunson & Park (2008). This prior is similar to the DDP in that it can be represented as an infinite sum of weights and point masses. The main difference is that the dependence on dose level in the KSBP occurs in the weights with common point masses. Under this type of prior, $G_{\mathcal{X}} = \sum_{l=1}^{\infty} \omega_{l\mathcal{X}} \delta_{\eta_l}$, the dose-response curve is given by $\Pr(y^* = 1; G_x) = \sum_{l=1}^{\infty} \omega_l(x)\pi(\eta_l)$. Hence, it can be readily shown that $\mathrm{E}\{\Pr(y^* = 1; G_x)\}$ is constant in x, irrespectively of the stochastic process used for the $\omega_{l\mathcal{X}} = \{\omega_l(x) : x \in \mathcal{X}\}$. This is an irremediable limitation for this type of prior specification, since the increasing trend in prior expectation is critical to anchor the estimation of the dose-response curve at unobserved toxin levels.

MCMC posterior simulation details

In the data sets we studied from the National Toxicology Program database, the dams are labeled and recorded in ascending numerical order across dose levels. Therefore, they can be associated as a response vector across the dose levels, and the replicated response vectors would then be considered to be exchangeable; see Fronczyk & Kottas (2014). Here, we assume the more natural ANOVA style formulation where dams are considered exchangeable both across and within dose levels. However, we note that for the data examples we have studied, inference results are very similar under the two versions of the hierarchical model formulation for the data.

To fix notation, let m_{ij} be the number of implants and R_{ij} the number of prenatal deaths for the jth dam at dose x_i, for $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$. Moreover, denote by $\boldsymbol{y}^*_{ij} = \{y^*_{ijk} : k = 1, \ldots, m_{ij} - R_{ij}\}$ and $\boldsymbol{u}^*_{ij} = \{u^*_{ijk} : k = 1, \ldots, m_{ij} - R_{ij}\}$ the corresponding pup specific binary malformation and continuous weight responses, respectively. In addition, $\boldsymbol{x} = (x_1, ..., x_N)$ denotes the vector of observed toxin levels, and $G_{\boldsymbol{x}}$ the as-

sociated mixing distribution which follows a $\mathrm{DP}(\alpha, G_{0\boldsymbol{x}})$ prior implied by the DDP prior for $G_{\mathcal{X}}$. Here, $G_{0\boldsymbol{x}}$ is given by a product of three N-variate normal distributions induced by the corresponding GPs used to define $G_{0\mathcal{X}}$. Finally, let

$$k(R, \boldsymbol{y}^*, \boldsymbol{u}^* \mid \gamma, \theta, \mu, \varphi) = \mathrm{Bin}(R; m, \pi(\gamma)) \prod_{k=1}^{m-R} \mathrm{Bern}(y_k^*; \pi(\theta))\mathrm{N}(u_k^*; \mu, \varphi),$$

denote the kernel of the DDP mixture model.

The mixture model for the data can be expressed in hierarchical form by introducing mixing parameters $(\gamma_{ij}(\boldsymbol{x}), \theta_{ij}(\boldsymbol{x}), \mu_{ij}(\boldsymbol{x}))$ for the jth observation $(R_{ij}, \boldsymbol{y}_{ij}^*, \boldsymbol{u}_{ij}^*)$ at dose x_i, where $\gamma_{ij}(\boldsymbol{x}) = (\gamma_{ij}(x_1), ..., \gamma_{ij}(x_N))$ (with the analogous notation for $\theta_{ij}(\boldsymbol{x})$ and $\mu_{ij}(\boldsymbol{x})$). Conditionally on $G_{\boldsymbol{x}}$, the $(\gamma_{ij}(\boldsymbol{x}), \theta_{ij}(\boldsymbol{x}), \mu_{ij}(\boldsymbol{x}))$ are independently distributed according to $G_{\boldsymbol{x}}$, and conditionally on the $(\gamma_{ij}(\boldsymbol{x}), \theta_{ij}(\boldsymbol{x}), \mu_{ij}(\boldsymbol{x}))$ and φ, the $(R_{ij}, \boldsymbol{y}_{ij}^*, \boldsymbol{u}_{ij}^*)$ are independently distributed according to $k(\cdot \mid \gamma_{ij}(x_i), \theta_{ij}(x_i), \mu_{ij}(x_i), \varphi)$, for $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$.

We proceed with MCMC posterior simulation via blocked Gibbs sampling (e.g., Ishwaran & James, 2001), which replaces the countable representation in Equation (1) of the paper with a finite truncation approximation. In particular, for the mixing distribution associated with the observed toxin levels we have $G_{\boldsymbol{x}} \approx \sum_{l=1}^{L} p_l \delta_{(V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x}))}$. Here, $p_l = \omega_l$, for $l = 1, ..., L-1$, and $p_L = 1 - \sum_{l=1}^{L-1} p_l$, and for $l = 1, ..., L$, $(V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x}))$ are independent from $G_{0\boldsymbol{x}}$. The truncation level L can be chosen using distributional properties for the weights arising from the DP stick-breaking structure. In particular, $\mathrm{E}(\sum_{l=1}^{L} \omega_l \mid \alpha) = 1 - \{\alpha/(\alpha+1)\}^L$, which can be averaged over the prior for α to estimate $\mathrm{E}(\sum_{l=1}^{L} \omega_l)$. For the analysis of the DYME data, we worked with $L = 50$, which under the gamma$(2, 1)$ prior for α, yields $\mathrm{E}(\sum_{l=1}^{50} \omega_l) = 0.99996$.

Now, consider configuration variable s_{ij} for the jth dam at dose x_i, such that $s_{ij} = l$, for $l = 1, \ldots, L$, if and only if $(\gamma_{ij}(\boldsymbol{x}), \theta_{ij}(\boldsymbol{x}), \mu_{ij}(\boldsymbol{x})) = (V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x}))$. With the

introduction of the s_{ij} the hierarchical model for the data becomes

$$\begin{aligned}
\{R_{ij}, \boldsymbol{y}^*_{ij}, \boldsymbol{u}^*_{ij}\} \mid \{(V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x})) : l = 1, \ldots, L\}, s_{ij}, \varphi &\overset{ind.}{\sim} \\
\mathrm{Bin}(R_{ij}; m_{ij}, \pi(V_{s_{ij}}(x_i))) \prod_{k=1}^{m_{ij}-R_{ij}} &\mathrm{Bern}(y^*_{ijk}; \pi(Z_{s_{ij}}(x_i)))\mathrm{N}(u^*_{ijk}; T_{s_{ij}}(x_i), \varphi) \\
s_{ij} \mid \boldsymbol{p} &\overset{i.i.d.}{\sim} \sum_{l=1}^{L} p_l \delta_l(s_{ij}), \;\; i = 1, \ldots, N; \;\; j = 1, \ldots, n_i \\
(V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x})) \mid \boldsymbol{\psi} &\overset{i.i.d.}{\sim} G_{0\boldsymbol{x}}, \;\; l = 1, \ldots, L
\end{aligned}$$

where the prior density for $\boldsymbol{p} = (p_1, ..., p_L)$ is given by a special case of the generalized Dirichlet distribution, $f(\boldsymbol{p} \mid \alpha) = \alpha^{L-1} p_L^{\alpha-1}(1-p_1)^{-1}(1-(p_1+p_2))^{-1} \times \cdots \times (1-\sum_{l=1}^{L-2} p_l)^{-1}$. The model is completed with hyperpriors (discussed in Section 3.1 of the paper) for φ, α, and $\boldsymbol{\psi}$.

Next, we provide details on sampling from the full conditional distributions required to implement the blocked Gibbs sampler. The key updates are for the mixing parameters $(V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x}))$. The corresponding full conditional distributions depend on whether l is associated with one of the active mixture components. Denote by $\{s^*_r : r = 1, \ldots, n^*\}$ the distinct values among the s_{ij}. If $l \notin \{s^*_r : r = 1, \ldots, n^*\}$, then $(V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x}))$ is drawn from the base distribution $G_{0\boldsymbol{x}}$ (given its currently imputed hyperparameters $\boldsymbol{\psi}$). When $l \in \{s^*_r : r = 1, \ldots, n^*\}$, the posterior full conditional for $V_l(\boldsymbol{x})$ is proportional to

$$\mathrm{N}_N(V_l(\boldsymbol{x}); \xi_0 \mathbf{1}_N + \xi_1 \boldsymbol{x}, \Lambda) \prod_{\{(i,j): s_{ij} = l\}} \mathrm{Bin}\left(R_{ij}; m_{ij}, \pi(V_l(x_i))\right)$$

where $\mathbf{1}_N$ is a vector of dimension N with all its elements equal to 1, and the covariance matrix Λ has elements $\tau^2 \exp\{-\rho|x_i - x_{i'}|\}$. Sampling from this full conditional was approached in several ways, including slice sampling and Metropolis-Hastings (M-H) random-walk updates with different choices for the proposal covariance matrix. The best

mixing and acceptance rates were obtained from a Gaussian proposal distribution with covariance matrix of the same form as the GP prior, that is, $a \exp\{-b|x_i - x_{i'}|\}$, where a and b are tuning parameters. The $Z_l(\boldsymbol{x})$ and $T_l(\boldsymbol{x})$ corresponding to active components are sampled in the same fashion. These M-H updates can be tuned to obtain sufficiently large acceptance rates; for instance, for the DYME data the acceptance rates ranged between 15% and 20%.

The full conditional for each s_{ij} is a discrete distribution on $\{1, ..., L\}$ with probabilities proportional to $p_l k(R_{ij}, \boldsymbol{y}^*_{ij}, \boldsymbol{u}^*_{ij} \mid V_l(x_i), Z_l(x_i), T_l(x_i), \varphi)$, for $l = 1, ..., L$. The updates for $\boldsymbol{p}$ and α are the same with the ones for a generic DP mixture model. Based on the inverse gamma prior for φ (with shape parameter $a_\varphi > 1$ and mean $b_\varphi/(a_\varphi - 1)$), its posterior full conditional is inverse-gamma with revised parameters $a_\varphi + 0.5 \sum_{i=1}^{N} \sum_{j=1}^{n_i} (m_{ij} - R_{ij})$ and $b_\varphi + 0.5 \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}-R_{ij}} (T_{s_{ij}}(x_i) - u^*_{ijk})^2$.

Regarding the parameters, $\boldsymbol{\psi} = (\xi_0, \xi_1, \tau^2, \rho, \beta_0, \beta_1, \sigma^2, \phi, \chi_0, \chi_1, \nu^2, \kappa)$, of the base distribution $G_{0\boldsymbol{x}}$, ξ_0, β_0 and χ_0 have normal posterior full conditional distributions, and τ^2, σ^2 and ν^2 have inverse gamma posterior full conditional distributions. We sample ξ_1, β_1 and χ_1 through random-walk M-H steps with normal proposal distributions on the log scale. To update the GP correlation parameters ρ, ϕ and κ, we have experimented with M-H steps, but ultimately found the most efficient approach to sample these parameters was through discretization of their underlying support induced by the uniform prior.

Finally, to extend the inference beyond the N observed dose levels, we can estimate the various risk assessment functionals at M new doses, $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_M)$, which may include values outside the range of the observed doses. Owing to the underlying DDP prior model for the mixing distributions, the only additional sampling needed is for the mixing parameters $(\tilde{V}_l(\tilde{\boldsymbol{x}}), \tilde{Z}_l(\tilde{\boldsymbol{x}}), \tilde{T}_l(\tilde{\boldsymbol{x}}))$ associated with the new dose levels. The product of GPs structure for the DDP base stochastic process implies an M-variate normal distribution

for $\tilde{V}_l(\tilde{\boldsymbol{x}})$ conditionally on $V_l(\boldsymbol{x})$, and analogously for $\tilde{Z}_l(\tilde{\boldsymbol{x}})$ conditionally on $Z_l(\boldsymbol{x})$, and for $\tilde{T}_l(\tilde{\boldsymbol{x}})$ conditionally on $T_l(\boldsymbol{x})$. Hence, given the posterior samples for $(V_l(\boldsymbol{x}), Z_l(\boldsymbol{x}), T_l(\boldsymbol{x}))$ and other model hyperparameters, we can readily sample the $(\tilde{V}_l(\tilde{\boldsymbol{x}}), \tilde{Z}_l(\tilde{\boldsymbol{x}}), \tilde{T}_l(\tilde{\boldsymbol{x}}))$, and consequently obtain inference for the various dose-response curves and response distributions at any desired grid over toxin levels.

Additional results from the DYME data analysis

Due to space restrictions, a full investigation of the probability mass functions is not found within the manuscript. Below, these inferences are given with discussion.

Also of interest is inference for the various response distributions. Figure 1 plots estimates for the probability mass functions corresponding to the number of non-viable fetuses given a specific number of implants. Figure 2 shows estimates for the probability mass functions of the number of malformations given a specified number of implants and associated number of non-viable fetuses. The DDP mixture model uncovers standard distributional shapes for most of the toxin levels, although there is evidence of skewness at $x = 250$ mg/kg. Finally, Figure 3 includes estimates for fetal weight densities. As expected, posterior uncertainty is larger at the unobserved dose level, $x = 175$ mg/kg. The spread of the densities is the same across toxin levels but the center shifts toward smaller fetal weight values under increasing dose values. Also noteworthy is the smooth evolution from left to right skewness in the densities as the toxin level increases.

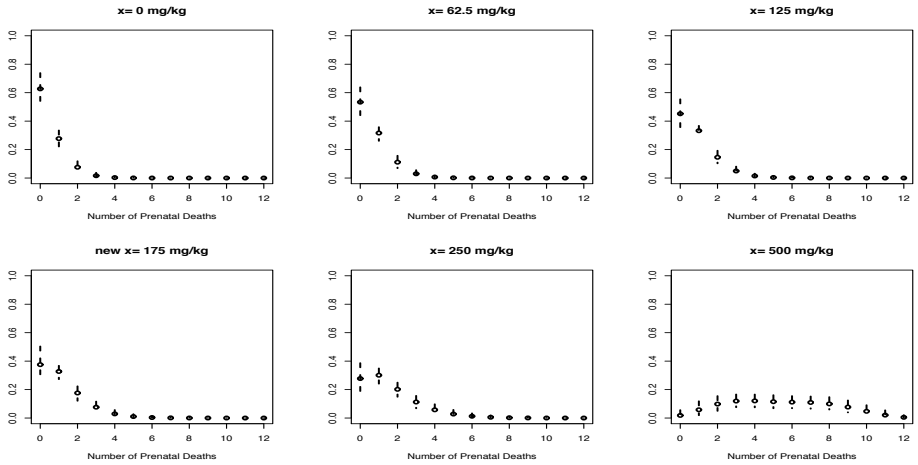

Figure 1: Posterior mean ("o") and 90% uncertainty bands (dashed lines) for the probability mass function associated with the number of non-viable fetuses given $m = 12$ implants, $f(R \mid m = 12, G_x)$. Results are shown for the five observed dose levels and for the new value of $x = 175$ mg/kg.

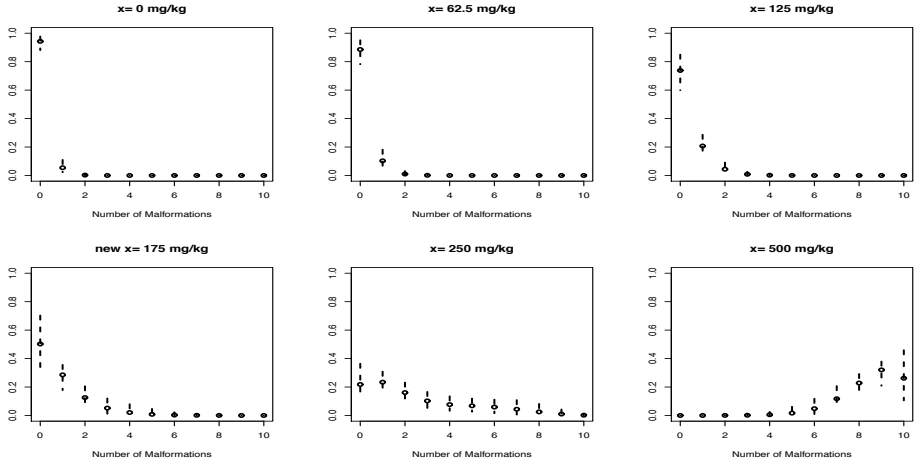

Figure 2: Posterior mean ("o") and 90% uncertainty bands (dashed lines) for the probability mass function associated with the number of malformations given $m = 12$ implants and $R = 2$ non-viable fetuses, $f(\sum_{k=1}^{m-R} y_k^* \mid m = 12, R = 2, G_x)$. Results are shown for the five observed dose levels and for the new value of $x = 175$ mg/kg.

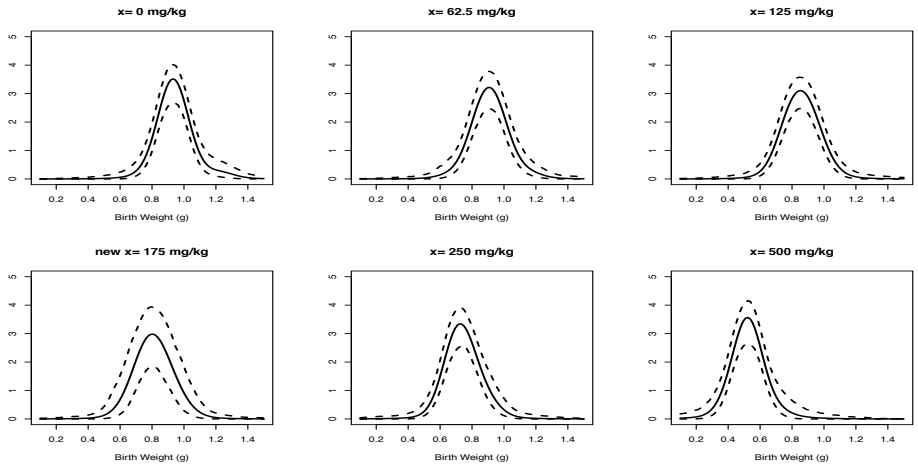

Figure 3: Posterior mean (solid line) and 90% uncertainty bands (dashed lines) for the probability density function for fetal weight, $f(u^* \mid m = 1, R = 0, G_x)$. Results are shown for the five observed dose levels and for the new value of $x = 175$ mg/kg.

References

Dunson, D. & Park, J. (2008). Kernel stick-breaking processes. *Biometrika* **95**, 307–323.

Fronczyk, K. & Kottas, A. (2014). A Bayesian nonparametric modeling framework for developmental toxicity studies (with discussion). *Journal of the American Statistical Association* **109**, 873–893.

Ishwaran, H. & James, L. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173.

Price, C., Kimmel, C., George, J. & Marr, M. (1987). The developmental toxicity of diethylene glycol dimethyl ether in mice. *Fundamentals of Applied Pharmacology* **8**, 115–126.