

Convex Mixture Regression for Quantitative Risk Assessment

Antonio Canale*, Daniele Durante[†] and David B. Dunson[‡]

May 9, 2018

Abstract

There is wide interest in studying how the distribution of a continuous response changes with a predictor. We are motivated by environmental applications in which the predictor is the dose of an exposure and the response is a health outcome. A main focus in these studies is inference on dose levels associated with a given increase in risk relative to a baseline. Popular methods either dichotomize the continuous response or focus on modeling changes with the dose in the expectation of the outcome. Such choices may lead to information loss and provide inaccurate inference on dose–response relationships. We instead propose a Bayesian convex mixture regression model that allows the entire distribution of the health outcome to be unknown and changing with the dose. To balance flexibility and parsimony, we rely on a mixture model for the density at the extreme doses, and express the conditional density at each intermediate dose via a convex combination of these extremal densities. This representation generalizes classical dose–response models for quantitative outcomes, and provides a more parsimonious, but still powerful, formulation compared to nonparametric methods, thereby improving interpretability and efficiency in inference on risk functions. A Markov chain Monte Carlo algorithm for posterior inference is developed, and the benefits of our methods are outlined in simulations, along with a study on the impact of DDT exposure on gestational age.

Keywords: Additional risk; Benchmark dose; Conditional density estimation; Convex density regression; Dose–response; Nonparametric density regression.

*Department of Statistical Sciences, University of Padua canale@stat.unipd.it

[†]Department of Decision Sciences, Bocconi University daniele.durante@unibocconi.it

[‡]Department of Statistical Science, Duke University dunson@duke.edu

1 INTRODUCTION

It is of substantial interest to study how the distribution of an outcome $y \in \mathcal{Y}$ varies with a predictor $x \in \mathcal{X} \subseteq \mathbb{R}$. Focusing on the case in which $\mathcal{Y} \subseteq \mathbb{R}$, with the response variable univariate and continuous, we let $f(y|x)$ denote the conditional density of y given x . Our focus is on environmental health studies in which x is the dose of a potentially adverse exposure, and y defines a health response. In such studies there is a wide interest in relating dose to the risk of an adverse health outcome, with such dose–response models forming the basis of quantitative risk assessment and regulatory guidelines on safe levels of exposure (e.g. Piegorsch and Bailer 2005, Chapter 4).

As a real–data illustrative application—which incorporates features common to different observational dose–response studies—we focus on relating DDE exposure in pregnant women to the risk of a premature delivery (Longnecker et al. 2001). The data set is obtained from a sub-study of the US Collaborative Perinatal Project (CPP), which was a large prospective study of US pregnant women and their children. In this sub-study, DDE, which is a persistent metabolite of the pesticide DDT, was measured in the maternal serum during the third trimester of pregnancy. DDE is lipophilic and is stored in fat tissues, so that women build up a body burden which could affect pregnancy.

Available risk assessment studies for these data rely on a Gaussian regression to infer changes in gestational age with DDE exposure, or leverage a logistic regression to study DDE effects on preterm birth, which coincides, typically, with a gestational age below 37 weeks. Although these models are useful, according to the histograms in Figure 1, the Gaussian assumption is unrealistic, whereas dichotomizing the data prior to inference leads to information loss (e.g. West and Kodell 1999). In fact, to obtain a more complete and accurate characterization of risk it is necessary to estimate how the whole distribution of the gestational age at delivery shifts with DDE. We address this goal via a novel Bayesian convex mixture regression (CoMiRe) model which allows the entire distribution of the gestational age to be unknown and changing with dose. The proposed formulation combines mixture representations and convex interpolations to balance flexibility and parsimony. This allows more accurate inference and quantitative risk assessment compared to Gaussian regression, and improves efficiency along with interpretability compared to fully nonparametric models.

1.1 Models and methods for quantitative risk assessment

In quantitative risk assessment one seeks to estimate the dose level associated with a given small increase in risk of an adverse health event relative to the background risk corresponding in general to zero dose. This estimated benchmark dose (BMD_q) is useful in supporting regulatory decisions, thus motivating inference on dose–response relationships to obtain the BMD_q —defined as the dose corresponding to a $100q\%$ risk of an adverse health event above

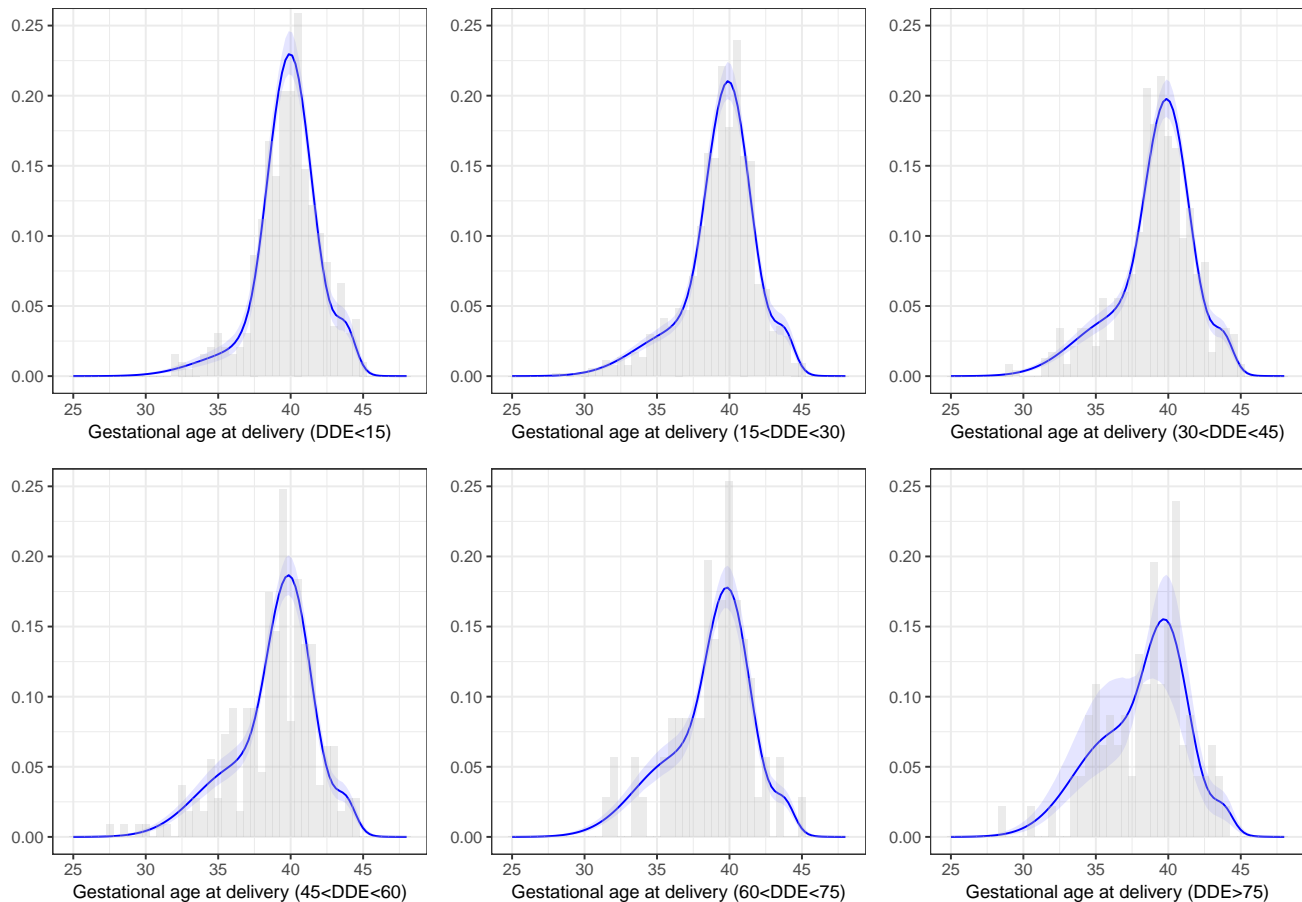


Figure 1: Histograms of the observed gestational age at delivery for selected dose intervals, along with the pointwise posterior mean (continuous blue lines), and 95% credible intervals (shaded blue areas) for $f(y | x)$ calculated in the central points (7.5, 22.5, 37.5, 52.5, 67.5, 125) of each dose interval, under the proposed convex mixture regression model. The figure appears in color in the electronic version of this article.

that for unexposed individuals.

Due to the abundance of dichotomous health data, the vast majority of the literature has focused on discrete responses (e.g. Piegorsch et al. 2014; Fronczyk and Kottas 2014). There has been instead less consideration of continuous outcomes. Earlier approaches dichotomize the continuous response via a clinical threshold and then apply models for binary data (e.g. Longnecker et al. 2001). From a statistical perspective, this leads to a loss of information, efficiency, and validity compared to dose–response studies on the original scale of the data (Crump 1995; West and Kodell 1999), thereby motivating a literature on quantitative risk assessment modeling the continuous response on its original scale. Popular formulations assume homoscedastic Gaussian responses with a dose–dependent expectation $\mu(x)$ commonly defined as $\mu(x) = \mu_0 + (\mu_\infty - \mu_0)\psi(x; \boldsymbol{\lambda})$ for every $x \in [0, \infty)$, with μ_0 and μ_∞ denoting the expectations of y at $x = 0$ and $x \rightarrow \infty$, respectively, whereas $\psi(x; \boldsymbol{\lambda})$ represents a monotone nondecreasing dose–response function having $\psi(0; \boldsymbol{\lambda}) = 0$, and $\psi(x; \boldsymbol{\lambda}) \rightarrow 1$ for $x \rightarrow \infty$ (Ritz et al. 2013, 2015). Other parametric models for $\mu(x)$ can be found in West and Kodell (1999); Kodell and West (1993); Piegorsch et al. (2005).

Once the expectation $\mu(x)$ has been estimated, the excess risk above the background is commonly measured via the additional risk function $R_A(x, a)$ (Kodell and West 1993), which is defined as

$$R_A(x, a) = \text{pr}(y \leq a \mid x) - \text{pr}(y \leq a \mid x = 0) = F_x(a) - F_0(a), \quad (1)$$

where $a \in \mathcal{Y}$ corresponds, in general, to a threshold of clinical interest, and $F_x(a) = F(a \mid x)$ is the conditional cumulative distribution function of the response at dose exposure x , computed in the threshold a . In the absence of relevant clinical guidelines for the cutoff value a , it is common practice to rely on tail values such as $a = \mu_0 - k\sigma$ with $k = 2$ or $k = 3$ (Kodell and West 1993; West and Kodell 1999; Piegorsch et al. 2005). Yu and Catalano (2005) consider instead quantile regression, which avoids specification of a cutoff and incorporates heteroscedasticity in Gaussian data. Based on (1), the BMD_q of interest for inference is the solution of the equation $R_A(x, a) = q$. Hence, differently from dichotomization strategies, the above methods leverage the entire information in the data to learn $F_x(a)$, and require a cutoff—e.g. $a = 37$ in our application—after the model has been estimated. According to (1), reliable inference on $R_A(x, a)$ and BMD_q requires accurate learning of $F_x(a)$, corresponding to an unknown continuous conditional distribution.

Although the aforementioned methods facilitate tractable inference on $F_x(a)$ and $R_A(x, a)$, the Gaussian assumption is difficult to justify in many toxicology studies. For example, as shown in Figure 1, this assumption is violated in our illustrative data set, in terms of heteroscedasticity, skewness, and multimodality. Due to this, alternative contributions consider mixtures of Gaussians, which improve flexibility while maintaining tractable inference in risk assessments. Refer to Razzaghi and Kodell (2000), He et al. (2010), and Hwang and Pen-

nell (2014) for relevant formulations. These models rely on a general specification for the conditional density $f_x(y) = f(y | x)$, via

$$f_x(y) = \sum_{h=1}^H \nu_h(x) \frac{1}{\sigma_h} \phi \left\{ \frac{y - \mu_h(x)}{\sigma_h} \right\}, \quad (2)$$

where $\nu_h(x)$ is a probability weight for the h -th mixture component varying with x , and $\phi\{\cdot\}$ is the standard Gaussian density.

In performing inference under the above mixture model, Razzaghi and Kodell (2000) consider a representation with $H = 2$ mixture components and fixed probability weights $\nu_h(x) = \nu_h$ for all $x \in \mathcal{X}$. This choice is motivated by the nonresponse phenomenon in dose–response studies (e.g. Good 1979), but $H = 2$ components may be insufficiently flexible in many situations. An alternative possibility to address this issue is to fix H at a conservative upper bound—potentially infinite—and allow the data to inform on the occupied number of mixture components under a Bayesian framework. One example corresponds to the class of ANOVA dependent Dirichlet process (DDP) mixture models developed by De Iorio et al. (2004). Although this formulation provides a more general alternative to a finite mixture model, there are still flexibility issues associated with the assumption of constant probability weights. This has motivated more flexible representations allowing the weights to change with x (e.g. Dunson and Park 2008; Rodriguez and Dunson 2011). Notable applications in quantitative risk assessment inspired by these models can be found in He et al. (2010), and Hwang and Pennell (2014). However, the increased flexibility of these methods comes at a cost in terms of interpretability and parsimony in characterizing functionals of interest. This may lead to a loss of efficiency, thereby increasing uncertainty in regulatory guidelines.

One possibility to improve upon these methods is to incorporate further structure, balancing interpretability, efficiency and flexibility. As described in Section 2, we address this goal via COMIRE. Our formulation interpolates two extremal densities $f_0(y)$ and $f_\infty(y)$ —defined via mixture models—through a single monotone increasing function $\beta(x) \in [0, 1]$, which induces a flexible, yet interpretable and parsimonious, characterization of $f_x(y)$. Although inference for the statistical model outlined in Section 2 can proceed under different paradigms—including the frequentist one—we focus here on a Bayesian implementation. This allows coherent uncertainty quantification and tractable inference on relevant quantities for quantitative risk assessment—such as the additional risk $R_A(x, a)$, the BMD_q , and more conservative benchmark doses BMDL_q based on lower credible bounds of the BMD_q (e.g. Crump 1995; Piegorsch et al. 2005). These quantities are non-linear functions of the parameters, thereby requiring asymptotic results and approximations for tractable frequentist inference. Under a Bayesian approach, the posterior for these quantities can be obtained via Monte Carlo methods exploiting the posterior samples of the parameters. Moreover, a Bayesian treatment allows formal incorporation of prior knowledge and inclusion of relevant restrictions. This Bayesian imple-

mentation is described in Section 3, and is compared against notable competitors on several simulation experiments in Section 4. The CPP application is explored in Section 5, whereas Section 6 contains concluding remarks and generalizations.

The code, data, and tutorial implementation of CoMiRe are available at github.com/tonycanale/CoMiRe. Additional results and analyses can be found also in the Supplementary Materials.

2 CONVEX MIXTURE REGRESSION

Let $y \in \mathcal{Y} \subseteq \mathbb{R}$. Our focus is on modeling the conditional density $y \mid x \sim f_x(y)$ via the mixture model

$$f_x(y) = f(y \mid x) = \int_{\Theta} K(y; \boldsymbol{\theta}) dP_x(\boldsymbol{\theta}), \quad (3)$$

where $K(y; \boldsymbol{\theta})$ is a kernel with support on \mathcal{Y} and parametrized by $\boldsymbol{\theta} \in \Theta$, whereas P_x is a mixing measure changing with x . Consistent with our focus, let $\mathcal{X} = \mathbb{R}^+$. To clarify how CoMiRe can be applied to many kernels, we avoid choosing a specific $K(y; \boldsymbol{\theta})$ for the moment.

Recalling Section 1.1, we look for a specification of P_x which maintains a degree of flexibility in modeling $f_x(y)$, while facilitating interpretable and efficient inference on relevant functionals for quantitative risk assessment. We address this goal by assuming P_x is a convex combination of the two mixing measures P_0 and P_∞ at $x = 0$ and $x \rightarrow \infty$, respectively, obtaining

$$P_x = \{1 - \beta(x)\}P_0 + \beta(x)P_\infty, \quad x > 0, \quad (4)$$

with $\beta(x)$ a bounded monotone nondecreasing interpolation function having $\beta(0) = 0$ and $\beta(x) \rightarrow 1$ as $x \rightarrow \infty$. Hence, changes in P_x with x are controlled by a nondecreasing function $\beta(x)$ inducing a continuous shift in P_x from P_0 to P_∞ , as x grows. There are several ways to define $\beta(x)$ under the conditions of monotonicity—i.e. $\beta'(x) \geq 0$ —and codomain constraints—i.e. $\beta(x) \in [0, 1]$. A simple strategy to address this goal is to consider the basis expansion

$$\beta(x) = \sum_{j=1}^J w_j \psi_j(x), \quad x > 0, \quad (5)$$

where $\psi_1(x), \dots, \psi_J(x)$ are monotone nondecreasing functions with $\psi_j(x) \in [0, 1]$ for all $j = 1, \dots, J$. With this choice, the constraints on $\beta(x)$ are translated into $0 \leq w_j \leq 1$ and $\sum_{j=1}^J w_j = 1$. As will be clarified in Section 3, this facilitates Bayesian inference. A tractable default choice for $\psi_1(x), \dots, \psi_J(x)$, which is provably flexible in shape-constrained inference, is the I-splines basis (Ramsay 1988). However, other specifications are possible, including general cumulative distribution functions or classical dose–response functions (Ritz et al. 2013, 2015).

To complete the specification of CoMiRe, we need a representation for the extreme mixing measures P_0 and P_∞ . In defining these quantities, we aim to preserve flexibility, while incorporating additional relevant structure available in quantitative risk assessment studies. In particular, the predictor is the dose of a chemical having increasingly adverse health effects at high values. Hence, as x grows, we expect the outcome to be centered on increasingly adverse health profiles, meaning that $E(y | x) = \mu(x)$ is monotone decreasing with x , when adverse health occurs for low values of y , as in our study. We include this adversity profile property and flexibility in $f_x(y)$, by letting

$$P_0 = \sum_{h=1}^H \nu_{0h} \delta_{\theta_{0h}}, \quad P_\infty = \delta_{\theta_\infty}, \quad (6)$$

with δ_θ being the Dirac's delta mass at point θ , H the total number of mixture components in $x = 0$, and appropriate restrictions for the atoms, so that the single mixture component at $x \rightarrow \infty$ is centered on a more adverse health profile than all the mixture components at $x = 0$, obtaining

$$\mu_\infty < \mu_{0h}, \quad \text{for every } h = 1, \dots, H. \quad (7)$$

In (7), μ_∞ denotes the expectation of the single mixture component at the extreme dose range, whereas μ_{0h} is the expectation of the mixture component h characterizing the density at the low end of the dose range, for every $h = 1, \dots, H$. Leveraging representations (3)–(6), equation (7) is sufficient to guarantee that $f_x(y)$ will shift to be centered on more adverse values of the health response as dose x increases. Indeed, under (3)–(6), we obtain

$$\begin{aligned} E(y | x) = \mu(x) &= \{1 - \beta(x)\}E(y | 0) + \beta(x)E(y | \infty) \\ &= \boldsymbol{\nu}_0^\top \boldsymbol{\mu}_0 + (\mu_\infty - \boldsymbol{\nu}_0^\top \boldsymbol{\mu}_0)\beta(x), \end{aligned} \quad (8)$$

where $\boldsymbol{\nu}_0 = (\nu_{01}, \dots, \nu_{0H})^\top$ and $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0H})^\top$ are the vectors containing the mixing weights and the expectations, respectively, of the components at the low end.

Although it is possible to relax (7) in various ways without affecting the monotonicity assumption for $\mu(x)$, assuming the component at the high end of the dose range to be centered on a more adverse health profile than all the components at the low end has appealing interpretation in dose–response studies. Indeed, equation (7) implies that the single component at the very high end of \mathcal{X} corresponds to the most adverse health profile. Under (4) and (6), this profile has probability zero and $\beta(x)$, when $x = 0$ and $x > 0$, respectively. This assumption interestingly incorporates common nonresponse effects in toxicology studies (Good 1979) by allowing the exposed individuals to have a different degree $\beta(x)$ of susceptibility to dose, which increases with x . However, differently from current formulations (e.g. Razzaghi and Kodell

2000), CoMiRe improves flexibility in modeling the nonresponders density $f_0(y)$ via a mixture model accounting for possible adverse effects of unobserved chemicals. Since there is less information in the data to learn $f_\infty(y)$, due to sparsity at high dose exposures, the responders density is defined via a single kernel. Although this choice seems restrictive, in practice $x \rightarrow \infty$ is never observed and we do not attempt inference on $f_x(y)$ for very large x . This leads to robustness to the assumptions in $x \rightarrow \infty$, with $f_x(y)$ fully flexible at low dose and shifting steadily towards being centered on more adverse values as x grows.

Generalizations to include more complex $f_\infty(y)$ are possible. We attempted this extension in initial analyses but observed no evident improvements. As we will outline in Sections 4 and 5, our basic CoMiRe leads to accurate results, and therefore we avoid complications affecting interpretation.

2.1 Interpretation and properties

Model (3)–(4) has several interpretations worth exploring. A fundamental one for quantitative risk assessment relates $\beta(x)$ to the additional risk function $R_A(x, a) = F_x(a) - F_0(a)$ via

$$\begin{aligned} R_A(x, a) &= \{1 - \beta(x)\}F_0(a) + \beta(x)F_\infty(a) - F_0(a) \\ &= \beta(x)\{F_\infty(a) - F_0(a)\} \\ &= \beta(x)R_A(\infty, a), \end{aligned} \tag{9}$$

where $F_0(a) = \text{pr}(y \leq a \mid 0)$, and $F_\infty(a) = \text{pr}(y \leq a \mid \infty)$. According to (9), the benefit associated with our representation, compared to the unstructured density regression models discussed in Section 1.1, is that the effect of x enters only via the flexible function $\beta(x)$ which is proportional to $R_A(x, a)$. Indeed, consistent with (9), the dose–response function $\beta(x)$ can be interestingly interpreted as a standardized additional risk at x with respect to an arbitrarily high dose. In fact, $\beta(x) = R_A(x, a)/R_A(\infty, a)$. This allows inference on the BMD_q , which is available as the solution of the equation $\beta(x) = q/R_A(\infty, a)$, for each $a \in \mathcal{Y}$ and benchmark risk q .

Besides the above interpretation, the formula for the conditional density $f_x(y)$ induced by model (3)–(4) can be intuitively seen also as the result of a process which travels in distribution between the starting location $f_0(y)$ and the ending one $f_\infty(y)$, as *time* x increases from 0 to ∞ , with $\beta(x)$ denoting the proportion of the path traveled at *time* x . This metaphor eases the interpretation of $\beta(x)$, and is formally supported by Theorem 1 in the Supplementary Materials.

To conclude the discussion on model interpretation and properties note that, consistent with (8), the conditional expectation $\mu(x)$ of the health outcome y induced by our CoMiRe coincides with the one characterizing popular dose–response models discussed in Section 1.1 (Ritz et al. 2013, 2015), after letting $\mu_0 = \boldsymbol{\nu}_0^\top \boldsymbol{\mu}_0$, $\omega_1 = 1$ and $\psi_1(x) = \psi(x; \boldsymbol{\lambda})$. At the

same time, it is also clear that, while relaxing the Gaussian assumption, the proposed basis expansion for $\beta(x)$ in (5) is more flexible compared to the parametric dose–response functions of routine use, thus generalizing the contributions of Wheeler and Bailer (2012) and Piegorsch et al. (2014).

3 BAYESIAN INFERENCE AND IMPLEMENTATION

Consistent with our application and the dose–response models outlined in Section 1.1, we focus on $y \in \mathcal{Y} \subseteq \mathbb{R}$, and let $K(y; \boldsymbol{\theta}) = \phi(y; \mu, \tau^{-1})$ be a Gaussian density with mean μ and variance $\sigma^2 = \tau^{-1}$, with $\boldsymbol{\theta} = (\mu, \tau)$. This choice also facilitates quantitative risk assessment, provided that, under the Gaussian assumption, $F_0(a) = \sum_{h=1}^H \nu_{0h} \Phi\{(a - \mu_{0h})/\sigma_{0h}\}$ and $F_\infty(a) = \Phi\{(a - \mu_\infty)/\sigma_\infty\}$.

Before describing the Markov chain Monte Carlo (MCMC) routine for posterior computation, we first include additional details on representations (3)–(6), and elicit the priors for each unknown quantity. Let us first introduce the augmented data (b_i, c_i, d_i) , for each $i = 1, \dots, n$, with $b_i \in \{1, \dots, J\}$ indicating the basis function associated with unit i in (5), $c_i \in \{1, \dots, H\}$ its mixture component at the low end of the dose range according to (6), and $d_i \in \{0, 1\}$ an indicator for the membership to one of the two extreme mixing measures in (4). Conditioned on these augmented data we can define the following hierarchical relations consistent with our model:

$$\begin{aligned} \{f_{x_i}(y) \mid d_i\} &= (1 - d_i)f_0(y) + d_i f_\infty(y), & d_i &\sim \text{Bern}\{\beta(x_i)\}, \\ \{\beta(x_i) \mid b_i = j\} &= \psi_j(x_i), & b_i &\sim \text{Cat}(w_1, \dots, w_J), \\ \{\boldsymbol{\theta}_i \mid c_i = h, d_i = 0\} &= \boldsymbol{\theta}_{0h}, & c_i &\sim \text{Cat}(\nu_{01}, \dots, \nu_{0H}), \end{aligned} \tag{10}$$

where $\text{Cat}(\rho_1, \dots, \rho_p)$ denotes a categorical variable with probabilities (ρ_1, \dots, ρ_p) , $\text{Bern}(\rho)$ denotes a Bernoulli distribution with success probability ρ , while $\boldsymbol{\theta}_{0h} = (\mu_{0h}, \tau_{0h})$. Marginalizing out the augmented data in (10), we obtain again (3), (5) and (6), under the convexity assumption in (4). This augmented form is useful to develop the MCMC.

To conclude our Bayesian specification, we need to elicit prior distributions for the model parameters, including the mixture weights in (6), the basis coefficients of $\beta(x)$ in (5), and each kernel-specific set of parameters. A natural prior for $\boldsymbol{\nu}_0 = (\nu_{01}, \dots, \nu_{0H})^\top$ at $x = 0$ is the Dirichlet distribution $\text{Dir}(\boldsymbol{\alpha})$ with vector of hyperparameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_H)$. According to Rousseau and Mengersen (2011), setting small values for the elements in $\boldsymbol{\alpha}$ provides a flexible prior for location-scale mixtures of Gaussians, which allows adaptive deletion of redundant components. Following a similar reasoning we define a Dirichlet prior for the coefficients of $\beta(x)$ in (5), so that $\boldsymbol{w} = (w_1, \dots, w_J)^\top \sim \text{Dir}(\boldsymbol{\eta})$, with small values for the hyperparameters in $\boldsymbol{\eta} = (\eta_1, \dots, \eta_J)$. Finally, for the atoms $\boldsymbol{\theta}_{0h} = (\mu_{0h}, \tau_{0h})$, $h = 1, \dots, H$, and $\boldsymbol{\theta}_\infty = (\mu_\infty, \tau_\infty)$

characterizing the Gaussian kernels, we choose independent gamma priors $\text{Ga}(a_\tau, b_\tau)$ for the precision parameters, and truncated Gaussian distributions for the locations μ_{0h} , $h = 1, \dots, H$ and μ_∞ , with mean $\mu \in \mathbb{R}$, variance $\kappa > 0$, and truncations meeting the adversity profile property in (7), $\mu_\infty < \min_h(\mu_{0h})$.

Under the above prior specification, samples from the posterior distribution can be obtained by sampling iteratively from the partially collapsed Gibbs sampler comprising the following steps.

1. Update b_i from the full conditional categorical random variable having probabilities

$$\text{pr}(b_i = j \mid -) \propto w_j [\{1 - \psi_j(x_i)\} f_0(y_i) + \psi_j(x_i) f_\infty(y_i)],$$

for every $j = 1, \dots, J$.

2. Update \mathbf{w} from the full conditional Dirichlet distribution

$$(\mathbf{w} \mid -) \sim \text{Dir}(\eta_1 + n_1, \dots, \eta_J + n_J),$$

where n_j is the number of subjects in which $b_i = j$, and update $\beta(x)$ by applying equation (5).

3. Update d_i from

$$(d_i \mid -) \sim \text{Bern} \left[\frac{\beta(x_i) f_\infty(y_i)}{\{1 - \beta(x_i)\} f_0(y_i) + \beta(x_i) f_\infty(y_i)} \right].$$

4. For individuals with $d_i = 0$, update c_i from the categorical variable having probabilities equal to

$$\text{pr}(c_i = h \mid -) \propto \nu_{0h} \phi(y_i; \mu_{0h}, \tau_{0h}^{-1}),$$

for every $h = 1, \dots, H$.

5. Update the mixture weights $\boldsymbol{\nu}_0$ characterizing P_0 from

$$(\boldsymbol{\nu}_0 \mid -) \sim \text{Dir}(\alpha_1 + n_1, \dots, \alpha_H + n_H),$$

where n_h is the number of units with $d_i = 0$, having $c_i = h$.

6. Update each $\boldsymbol{\theta}_{0h} = (\mu_{0h}, \tau_{0h})$, $h = 1, \dots, H$ under the restriction $\mu_\infty < \min_h(\mu_{0h})$, from

$$\begin{aligned} (\mu_{0h} \mid -) &\sim \text{N}(\hat{\mu}_h, \hat{\sigma}_h^2, \mu_\infty, +\infty), \\ (\tau_{0h} \mid -) &\sim \text{Ga}(\hat{a}_h, \hat{b}_h), \end{aligned}$$

where $\text{N}(\mu, \sigma^2, a, b)$ denotes a Gaussian with mean μ , variance σ^2 , and truncated to the interval (a, b) . In the above equation, $\hat{a}_h = a_\tau + n_h/2$, $\hat{b}_h = b_\tau + \sum_{i=1}^n \mathbb{I}(d_i = 0, c_i = h)(y_i - \mu_{0h})^2/2$, $\hat{\sigma}_h^2 = (\kappa^{-1} + n_h \tau_{0h})^{-1}$, $\hat{\mu}_h = \hat{\sigma}_h^2(\kappa^{-1} \mu + n_h \tau_{0h} \bar{y}_h)$, and \bar{y}_h is the mean of y for the units with $d_i = 0$ and $c_i = h$. The term $\mathbb{I}(\cdot)$ denotes, instead, the indicator function.

7. Finally, update $\boldsymbol{\theta}_\infty = (\mu_\infty, \tau_\infty)$ under the restriction $\mu_\infty < \min_h(\mu_{0h})$, from

$$\begin{aligned}(\mu_\infty | -) &\sim \text{N}(\hat{\mu}_\infty, \hat{\sigma}_\infty^2, -\infty, \min_h(\mu_{0h})), \\(\tau_\infty | -) &\sim \text{Ga}(\hat{a}_\infty, \hat{b}_\infty),\end{aligned}$$

where $\hat{a}_\infty = a_\tau + n_\infty/2$, $\hat{b}_\infty = b_\tau + \sum_{i=1}^n d_i(y_i - \mu_\infty)^2/2$, $\hat{\sigma}_\infty^2 = (\kappa^{-1} + n_\infty\tau_\infty)^{-1}$, $\hat{\mu}_\infty = \hat{\sigma}_\infty^2(\kappa^{-1}\mu + n_\infty\tau_\infty\bar{y}_\infty)$, $\bar{y}_\infty = \sum_{i=1}^n d_i y_i/n_\infty$, and $n_\infty = \sum_{i=1}^n d_i$.

This routine can be easily generalized to incorporate mixture models also for $f_\infty(y)$ simply modifying step (7), and adding a step similar to (4), to update the weights also at the high end of \mathcal{X} .

3.1 Goodness-of-fit via posterior predictive checks

To assess whether the restrictions of CoMiRe provide a good fit to the observed data under the Bayesian framework, we perform posterior predictive checks (Gelman et al. 2014) comparing relevant summary statistics computed from the observed data with their corresponding posterior predictive densities. Under CoMiRe, the posterior predictive density $f_x(y) | (y_1, x_1), \dots, (y_n, x_n)$ from which the response data can be simulated is defined as

$$\int \left[\{1 - \beta(x)\} \sum_{h=1}^H \nu_{0h} \tau_{0h} \phi\{\tau_{0h}(y - \mu_{0h})\} + \beta(x) \tau_\infty \phi\{\tau_\infty(y - \mu_\infty)\} \right] d\Pi(\boldsymbol{\nu}_0, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty, \boldsymbol{w} | \text{data}), \quad (11)$$

where $\Pi(\boldsymbol{\nu}_0, \boldsymbol{\theta}_0, \boldsymbol{\theta}_\infty, \boldsymbol{w} | \text{data})$ denotes the joint posterior distribution of the model parameters.

Although (11) is not analytically available, it is straightforward to simulate outcome data y from the posterior predictive distribution in (11) by leveraging the MCMC samples of the parameters and the hierarchical representation of the model in (10). Once these simulated data sets are available, different summary statistics of interest can be computed, thus obtaining samples from the corresponding posterior predictive distribution. If CoMiRe is not sufficiently flexible for a specific study, we expect the observed summary statistics to fall in the tails of their associated posterior predictive density. Since our main interest is in quantitative risk assessment, we focus on comparing a smoothed empirical estimate of $\text{pr}(y \leq a | x) = F_x(a)$ computed from the dichotomized sample data $\{\mathbb{I}(y_1 \leq a), x_1\}, \dots, \{\mathbb{I}(y_n \leq a), x_n\}$, with its posterior predictive density obtained from the replicated data sets generated from (11). Although other statistics can be considered, according to (1), $F_x(a)$ is the key function underlying $R_A(x, a)$, thus requiring careful model checking.

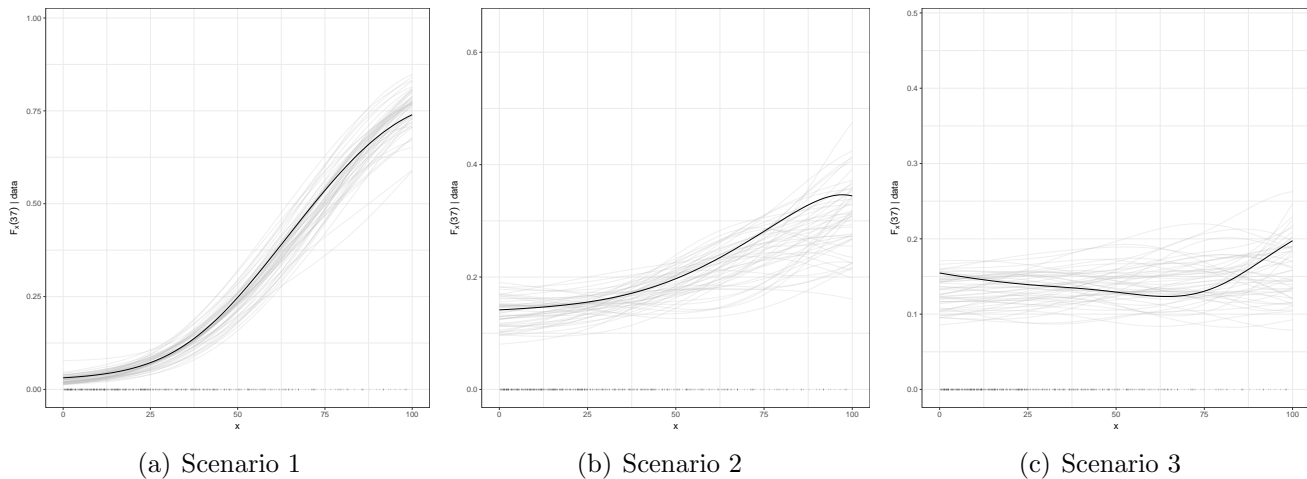


Figure 2: Goodness-of-fit assessments for $n = 500$ in Scenario 1 (a), 2 (b), and 3 (c). Smoothed empirical estimate of $F_x(37) = \text{pr}(y \leq 37 | x)$ computed from the observed data (black line), and from 50 data sets simulated from the posterior predictive distribution induced by CoMiRe (grey lines). In the x axis we also report the simulated dose exposures.

4 SIMULATION EXPERIMENTS

To study the empirical performance of CoMiRe in providing inference for $R_A(x, a)$ under several data generative processes, we consider three simulations, covering correctly specified models and model misspecification scenarios. To carefully assess performance in reasonable applied settings, these scenarios are defined to incorporate relevant aspects of the motivating CPP application and other related observational studies. These assessments involve two different sample sizes: $n = 2000$ to mimic the CPP application, and $n = 500$ to evaluate performance in smaller studies.

A common feature in observational dose–response studies is that the data are increasingly sparse as the dose grows, with most of the individuals having reasonably low exposures to adverse chemicals. We incorporate this feature by generating the doses $x_i, i = 1, \dots, n$, from a gamma distribution with a heavy right tail. These simulated doses range from 0 to about 132, and have empirical quartiles 15.4, 24.9, 39.8 for the scenarios with $n = 2000$, and 11.2, 25.0, 48.2, for those with $n = 500$. For each unit, y_i is generated under three different mechanisms which maintain a similar range and structure to the one characterizing the observed gestational age in the CPP data.

Scenario 1 covers the correctly specified setting and assumes the proposed CoMiRe model

where the extreme density at $x = 0$ is a mixture of three Gaussian kernels having $\boldsymbol{\mu}_0 = (37, 39, 40)^\top$, $\boldsymbol{\nu}_0 = (0.05, 0.15, 0.80)^\top$, and unit variances, whereas the density at $x \rightarrow \infty$ is a single Gaussian kernel with $\mu_\infty = 36$ and $\sigma_\infty^2 = 1$. The $\beta(x)$ function is assumed to be the cumulative distribution function of a gamma random variable with shape 6 and rate 0.1. Focusing on performance under model misspecification, we generate the data from a dependent process but without CoMiRe specification in Scenario 2, and under an independent process in Scenario 3. Specifically, Scenario 2 assumes a mixture of Gaussians with unit variances and locations changing with x . The mixing weights are $(0.10, 0.25, 0.65)$, whereas the locations are $\mu_1(x) = -x/300 + 35.5$, $\mu_2(x) = -x/50 + 38.5$, and $\mu_3(x) = -x/75 + 40.5$, respectively. Scenario 3 is instead a mixture of Gaussians with weights equal to $(0.25, 0.25, 0.50)$, locations $(37, 39, 41)$, and unit variances.

In performing posterior inference, we set the hyperparameters following the guidelines in Section 3. In particular we fix $H = 10$ for the number of mixture components at zero dose, and set small values $\alpha_1 = \dots = \alpha_H = 1/H$ in the Dirichlet prior for $\boldsymbol{\nu}_0$ to allow adaptive shrinkage. The function $\beta(x)$ in (5) is instead defined via cubic I-splines with 7 equally-spaced inner knots and 2 more at the extremes of the observed dose range—i.e. 0 and 132. The knot at $x = 0$ incorporates the restriction $\beta(0) = 0$, whereas the one at 132 introduces a zero constant function which allows asymptotes in $\beta(x)$ at extreme doses. This choice provides $J = 10$ bases. By considering small values $\eta_1 = \dots = \eta_J = 1/J$ for the hyperparameters in the Dirichlet prior for the I-splines coefficients \boldsymbol{w} , we facilitate automatic learning of which bases are required to flexibly characterize the dose–response function $\beta(x)$. Finally, we center the mean parameters in the Gaussian kernels on the empirical mean of the observed responses $\mu = \bar{y}$, and allow moderate variations in the expectation of the different components by setting $\kappa = 10$. For the hyperparameters in the component–specific variances we set instead $a_\tau = b_\tau = 2$ to allow a moderate variability within each Gaussian kernel. To assess sensitivity we maintain the same default hyperparameters settings in all the simulations.

Posterior inference relies on 5000 Gibbs iterations after a burn-in of 2000, and thinning the chains every 5 samples. These settings were sufficient to obtain convergence and good mixing—based on the traceplots and the Geweke’s diagnostic. The computations for each simulation took $\approx 2'$ and $\approx 6'$ for $n = 500$ and $n = 2000$, respectively, under an R implementation on a MacBook Pro with 2.8 GHz Intel i7 processor and 16 GB RAM. The following discussions are related to the case $n = 500$. The results for $n = 2000$ are similar and reported in the Supplementary Materials. It shall be also noticed that the following figures, including those in Section 5 and in the Supplementary Materials, evaluate performance in a large dose range avoiding excessively high exposures, which are unlikely to occur in the population and are not of direct interest in quantitative risk assessments.

We study performance with a particular focus on recovering the true additional risk $R_A(x, a)$. However, before focusing on $R_A(x, a)$, we first check model adequacy in Figure

2, following the guidelines provided in Section 3.1. As is clear from Figure 2, our approach provides adequate fit to the smoothed empirical estimate of $F_x(37) = \text{pr}(y \leq 37 | x)$ computed from the observed data, thereby motivating inference on the additional risk $R_A(x, a)$, with $a = 37$. Note that these plots can detect lack of fit—for example, if the dose–response relation is non–monotone, or if there are evident differences between the generative process of the observed data and the deflation–inflation mechanism of CoMiRe. In these situations, relaxing CoMiRe restrictions under more flexible models is important to avoid bias. We shall however emphasize that non–monotonicity of dose–response across the range of exposures faced by humans in usual settings is rare. The most common type of non–monotonicity in these studies corresponds to down–turns at high doses—typically well beyond normal ranges of human exposures. Similarly, evident departures from the CoMiRe assumptions, although possible, are not frequent in observational dose–response studies. Hence, we focus on monotone relations, and evaluate robustness to misspecification in reasonable scenarios.

To highlight the benefits of CoMiRe in providing inference on the additional risk, we compare performance with the relevant competitors discussed in Section 1.1. In particular, we consider a classical Gaussian regression with a Weibull dose–response function for $\mu(x)$ (Ritz et al. 2013, 2015), a more general formulation of the model in Razzaghi and Kodell (2000), via ANOVA–DDP with fixed weights (De Iorio et al. 2004), and, finally, a fully flexible dependent mixture of Gaussians (F–DMIX) including changes also in the mixing probabilities as in He et al. (2010). These models are estimated with the R packages `drc`, `DPpackage`, and `LSBP`, respectively. `LSBP` is a recent implementation of the probit stick–breaking by Rodriguez and Dunson (2011), using the logistic link to improve computational performance without affecting flexibility (Rigon and Durante 2017). In performing posterior inference under the ANOVA–DDP and the F–DMIX, we consider routine implementations via linear functions of the predictor, and set the hyperparameters using default choices which induce a comparable prior uncertainty to CoMiRe. Figure 3 summarizes the posterior mean function and pointwise 95% posterior credible bands for $R_A(x, 37)$ obtained under CoMiRe, ANOVA–DDP and F–DMIX, in the three scenarios. Results under the Weibull dose–response function were substantially inferior, and therefore are not reported here.

The results in Figure 3 confirm our discussion in Section 1.1 on the different approaches to inference in quantitative risk assessment. In particular, CoMiRe allows improved learning of $R_A(x, 37)$ in the correctly specified scenario, and has comparable performance to the more flexible ANOVA–DDP and F–DMIX, in misspecified settings. As it can be noticed in the lower panels of Figure 3, the estimation of $R_A(x, 37)$ is particularly precise at low–dose exposures x , where quantitative risk assessments typically focus. The reduced performance at high–dose exposures is mainly due to sparsity in the data, thereby providing reduced information to effectively estimate the parameters characterizing P_∞ . This result is also evident in Figure 2. The ANOVA–DDP provides accurate inference on $R_A(x, 37)$ in Scenarios 2 and 3, where this

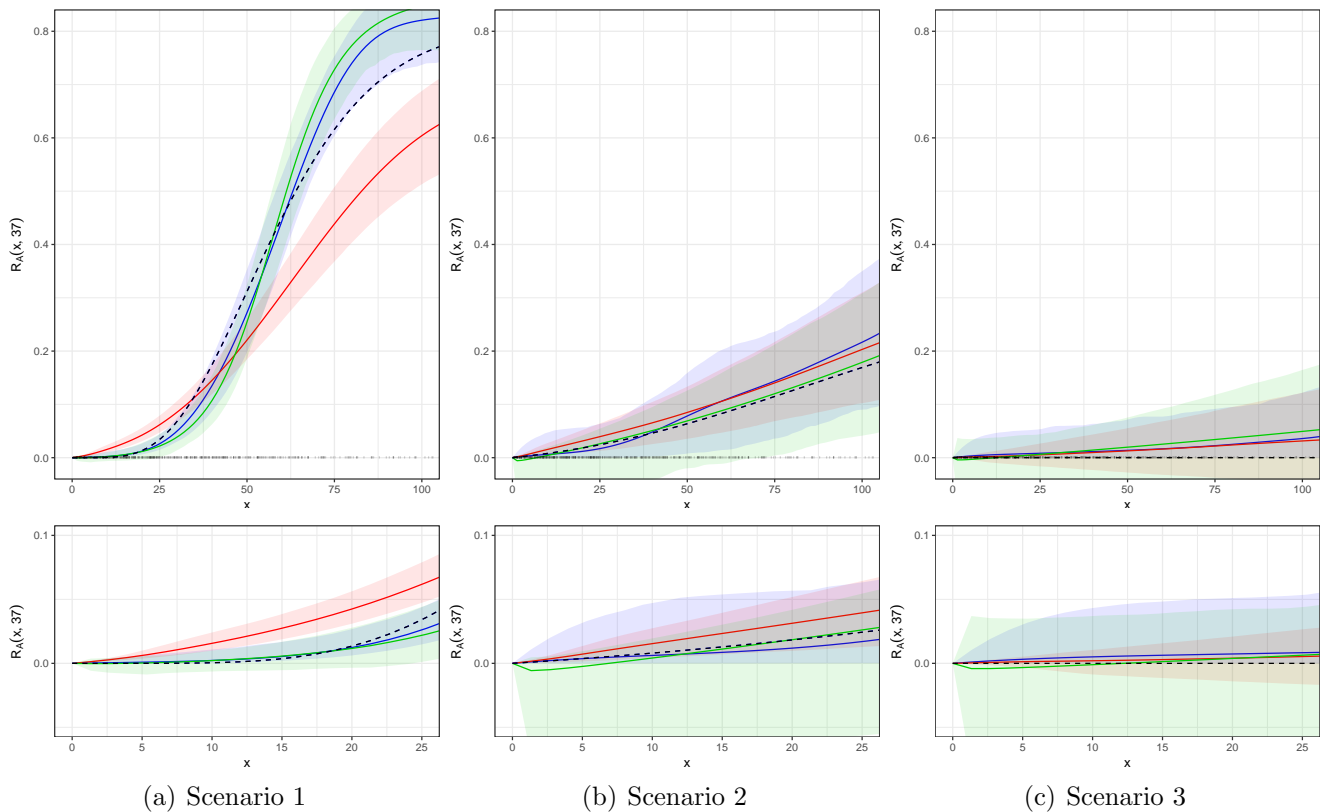


Figure 3: Inference on the additional risk function for $n = 500$ in the three scenarios. The dashed lines represent the true additional risk function $R_A(x, 37)$, whereas the red, green, and blue continuous lines denote the posterior mean of $R_A(x, 37)$ under ANOVA-DDP, F-DMIX, and CoMiRE, respectively. The shaded areas represent the pointwise 95% posterior credible bands. In the x axis we report the simulated dose exposures. Lower panels provide a zoom on the range of the additional risk typically considered in benchmark dose analysis.

model is correctly specified, but induces a notable bias in Scenario 1, possibly due to the assumption of constant mixing weights. When relaxing this restriction under F-DMIX, the bias is reduced in Scenario 1. However, consistent with the discussion in Section 1.1, the increased flexibility associated with F-DMIX requires more parameters, thereby reducing efficiency. This is evident in all the three scenarios, with F-DMIX having higher posterior uncertainty than CoMiRe, without substantially improving performance in estimating $R_A(x, 37)$. According to the lower panels in Figure 3, this reduced efficiency is evident for the common range $(0, 0.1]$ of the additional risk to be considered in benchmark dose analysis. These conclusions did not change when considering other thresholds a different from 37.

Similar results also are discussed for $n = 2000$ in the Supplementary Materials. In this case the higher sample size further reduces bias and posterior uncertainty—as expected.

5 ANALYSIS OF THE CPP DATA

We conclude by applying CoMiRe to infer changes in the gestational weeks at delivery with DDE, as discussed in Section 1. In particular, our focus is on $n = 2312$ women with gestational ages less or equal than 45 weeks, since higher values are clear measurement errors. Posterior inference is performed with the default hyperparameters and MCMC settings carefully described in the tutorial implementation at github.com/tonycanale/CoMiRe. Also in this case we obtain convergence and good mixing, with a runtime of $\approx 8'$.

As outlined in Figures 1 and 4, CoMiRe provides an adequate fit to the observed data and the associated functionals of interest, thus motivating inference and quantitative risk assessments. According to Figure 1, the conditional density of the gestational age at delivery is far from being Gaussian and displays variability, skewness and multimodality patterns changing with DDE. Indeed, at low-dose exposures most of the probability mass is concentrated around normal pregnancies, with the posterior mean and the 95% credible interval for $\mu(0)$ being 40.20 and (40.01, 40.34). This is in line with results on normal gestational ages measured via last menstrual periods, as in our application (Longnecker et al. 2001). In addition to these findings, we also notice mild negative skewness, with a minor mode closer to overdue pregnancies, and a tail towards preterm deliveries. As DDE exposure grows, the negative skewness is still maintained, but the tail towards preterm deliveries increasingly inflates until characterizing a more evident mode centered on μ_∞ . The posterior mean and the 95% credible interval for the latter quantity are 36.22 and (35.36, 36.94), respectively, thus providing evidence of preterm pregnancy profiles at high-dose exposures.

These results clearly show that the negative association of DDE with gestational age are mostly on the shape, rather than on the trend. In particular, recalling the above discussion, both adverse and non-adverse profiles are found across the predictor space, including at no dose and high-dose exposures. What changes with DDE is the degree $\beta(x)$ of susceptibility to the

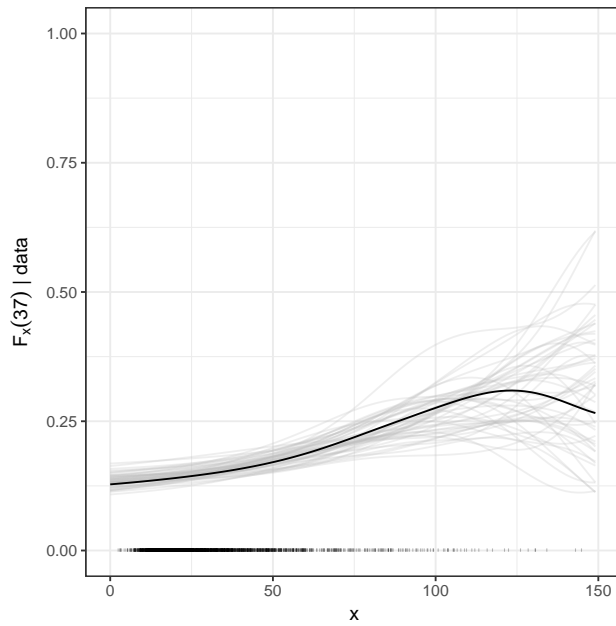


Figure 4: Goodness-of-fit assessment in the application. Smoothed empirical estimate of $F_x(37) = \text{pr}(y \leq 37 | x)$ computed from the observed data (black line), and from 50 data sets simulated from the posterior predictive distribution induced by CoMiRe (grey lines). In the x axis we report the observed exposures.

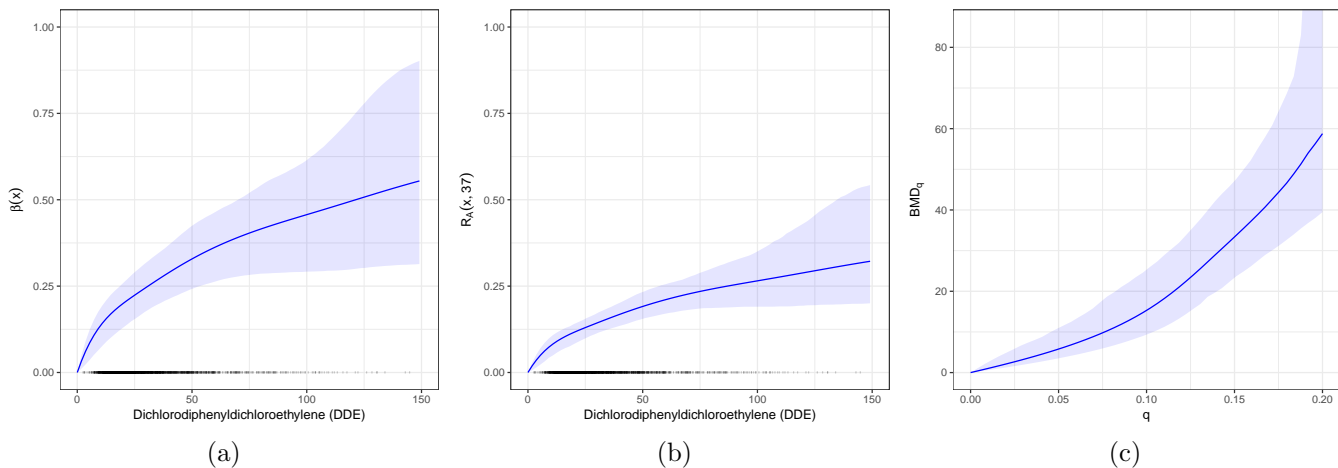


Figure 5: Posterior mean (solid lines) and pointwise 95% credible bands (shaded areas) for (a) $\beta(x)$, (b) $R_A(x, 37)$ and (c) the related BMD_q . In the x axis in (a) and (b) we report the observed exposures.

adverse effects of this chemical. In particular, the posterior mean and the 95% credible bands of the $\beta(x)$ function in panel (a) of Figure 5, show a notable increment in the probability of the most adverse health profile at low-dose exposures. For instance, about 20% of the women with a low DDE concentration $\approx 20\mu g/l$ are expected to have a gestational age comparable to women who had high DDE exposures, with this profile being indicative of preterm pregnancies. Consistent with the properties of `comiRe` discussed in Section 2.1, this percentage explicitly measures in which proportion the additional risk at a given DDE exposure relates to the one associated with the worst health profile observed at an arbitrarily high dose without the need to specify any threshold a for the negative health event.

For benchmark dose analyses, we consider the standard preterm threshold $a = 37$. Consistent with Section 2.1, rescaling $\beta(x)$ by $R_A(\infty, 37)$ provides the additional risk function $R_A(x, 37)$ —reported in panel (b) of Figure 5—inheriting the notable increment in the additional risk at low-dose exposures, which suggests conservative benchmark doses. The posterior mean and the 95% credible bands for the BMD_q —expressed as a function of q —confirm these findings. In particular, as shown in panel (c) of Figure 5, the BMD_q evolves on low values, especially for the range of benchmark risks $q \in \{0.01, 0.05, 0.10\}$ of interest in toxicology (e.g. Piegorsch et al. 2005), thus implying the conservative policies on DDE exposures in Table 1. Motivated by interest in more conservative benchmark doses $BMDL_q$ relying on a lower confidence bound of the BMD_q , Table 1 provides also the $BMDL_q$. Under our Bayesian approach to

	0.01	0.05	0.10
BMD _q	1.03 [0.61, 2.62]	5.79 [3.50, 11.00]	15.29 [9.33, 25.41]
BMDL _q	0.64	3.70	9.85

Table 1: For the three typical values of q , posterior mean and 95% credible intervals of the BMD_q, along with the BMDL_q, characterizing the 5% quantile of the posterior distribution of the BMD_q.

inference, such exposures can be obtained as the 5% lower quantile of the posterior distribution of the BMD_q, instead of requiring frequentist asymptotic approximations.

We also performed inference varying H to $H = 5$ and $H = 15$, but observed no relevant differences in inference on $R_A(x, 37)$, BMD_q, and BMDL_q.

6 CONCLUSION AND EXTENSIONS

Motivated by quantitative risk assessment, we proposed a class of CoMiReS balancing flexibility and parsimony in modeling conditional densities. Although the focus is on situations in which the range of dose exposures is unbounded, CoMiRe can be applied also in studies when x represents a concentration which is bounded above by x_{\max} . In such settings, one appealing possibility is to focus on the rescaled predictor $\bar{x} = x/x_{\max} \in [0, 1]$ and define each $\psi_j(\bar{x})$ in (5) via integrated Bernstein polynomials—i.e. $\text{Beta}(j, 2^{\bar{J}} - j + 1)$ —with $\bar{J} = \log_2\{J\}$. CoMiRe can be also applied to binary health responses. In such settings it suffices to consider Bernoulli kernels $K(y; \boldsymbol{\theta}) = \pi^y(1 - \pi)^{1-y}$, and let the extreme measures be $P_0 = \delta_{\pi_0}$ and $P_\infty = \delta_{\pi_\infty}$, to obtain a fully flexible specification.

Consistent with our motivating application, we focus on the case in which lower outcome values y are associated with adverse health. It is however straightforward to adapt the proposed model to scenarios in which higher values of y are more adverse—e.g. blood cholesterol level. It is also possible to control for the effect of additional covariates \mathbf{z} , while maintaining the adversity profile property in x . This can be done by allowing the location of each mixture component to change also as a function $g(\mathbf{z})$ of the additional covariates \mathbf{z} . Using the same $g(\mathbf{z})$ within each component, the adversity profile property in x is maintained, and the additional risk in x , for any \mathbf{z} , is $R_A(x, a; \mathbf{z}) = \beta(x)R_A\{\infty, a; g(\mathbf{z})\}$, with $R_A\{\infty, a; g(\mathbf{z})\}$ the additional risk at (∞, \mathbf{z}) . Note also that, although we focus on observational data, CoMiRe can be also applied in experimental studies where units are exposed to pre-specified doses. In such settings, CoMiRe will efficiently interpolate between the fixed doses. In contrast, fully unstructured models, that attempt to infer the continuous dose-response relation from limited dose groups, will tend to have wide uncertainty across the gaps between the pre-specified doses.

It is also interesting to incorporate U-shape behaviors for $\mu(x)$ at low dose exposures,

which may occur in toxicology studies (e.g. Calabrese and Baldwin 2001). One possibility to incorporate this property within CoMiRE is to introduce an intermediate mixing distribution P_u at the central value x_u corresponding to the minimum of the U-shape. Under appropriate adversity profile restrictions for the atoms in P_0 , P_u and P_∞ , the U-shape can be incorporated, including uncertainty in x_u .

A SUPPLEMENTARY MATERIAL

Formal Interpretation of the $\beta(x)$ Function

According to Section 2.1 of the paper, CoMiRE has several interpretations. Here, we formalize the metaphor interpreting the conditional density $f_x(y)$ as the result of a process which travels in distribution between the starting location $f_0(y)$ and the ending location $f_\infty(y)$, as *time* x increases from 0 to ∞ . Theorem 1 clarifies the role of the $\beta(x)$ function in measuring the proportion of the path travelled at *time* x .

Theorem 1. *Let $F_0(y)$, $F_\infty(y)$, and $F_x(y)$ characterize the probability distribution functions inducing the density functions $f_0(y)$, $f_\infty(y)$ and $f_x(y)$, respectively, in equations (3)–(4), and let*

$$d_{\text{TV}}\{F_1(y), F_2(y)\} = \frac{1}{2} \int |f_1(y) - f_2(y)| dy$$

be the total variation metric between the generic probability distribution functions $F_1(y)$ and $F_2(y)$ inducing the densities $f_1(y)$ and $f_2(y)$. Then

$$\beta(x) = d_{\text{TV}}\{F_x(y), F_0(y)\} / d_{\text{TV}}\{F_\infty(y), F_0(y)\}.$$

Proof. Theorem 1, can be easily derived after noticing that

$$\begin{aligned} \int |f_x(y) - f_0(y)| dy &= \int |\{1 - \beta(x)\}f_0(y) + \beta(x)f_\infty(y) - f_0(y)| dy \\ &= \beta(x) \int |f_\infty(y) - f_0(y)| dy = 2\beta(x) d_{\text{TV}}\{F_\infty(y), F_0(y)\}, \end{aligned}$$

and therefore $\beta(x) = d_{\text{TV}}\{F_x(y), F_0(y)\} / d_{\text{TV}}\{F_\infty(y), F_0(y)\}$, thus proving Theorem 1. \square

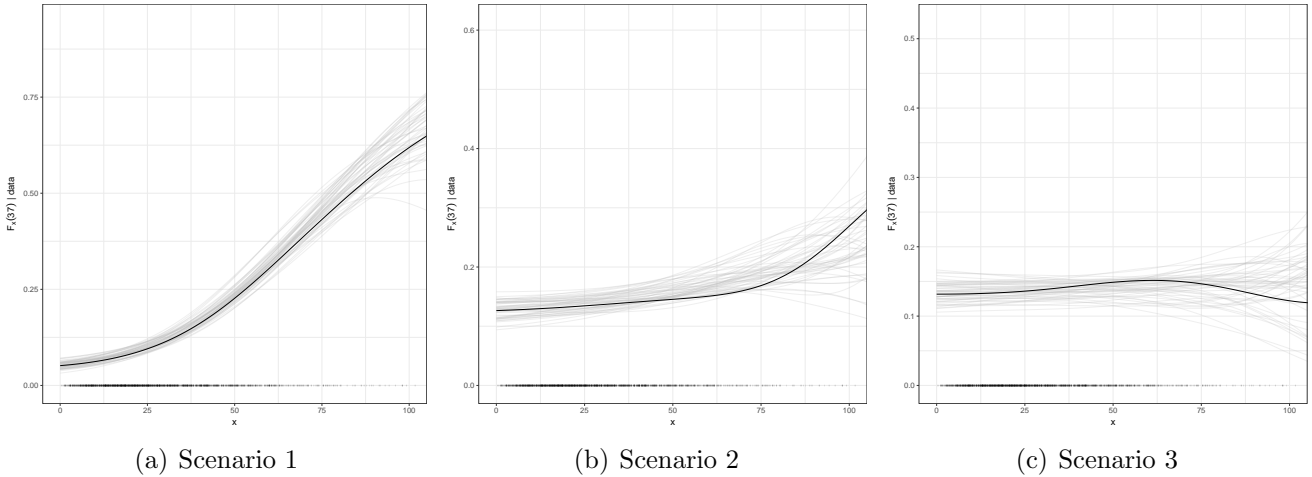


Figure 6: Goodness-of-fit assessments for $n = 2000$ in Scenario 1 (a), 2 (b), and 3 (c). Smoothed empirical estimate of $F_x(37) = \text{pr}(y \leq 37 | x)$ computed from the observed data (black line), and from 50 data sets simulated from the posterior predictive distribution induced by CoMiRe (grey lines). In the x axis we also report the simulated dose exposures.

Additional Simulation Results

Here, we reproduce the performance assessments of Section 4 in the paper under the same scenarios, simulation settings, prior specifications, and competing methods, but focusing on a larger sample involving $n = 2000$ units, instead of $n = 500$. This situation mimics the sample size available in the CPP application. Refer to Section 4 for details on the three scenarios, prior specifications, and competing methods.

Consistent with the performance assessments in Section 4, we first check model adequacy in Figure 6 and then focus on evaluating inference for $R_A(x, a)$ in Figure 7. According to Figure 6, also for $n = 2000$, our approach provides adequate fit to the smoothed empirical estimate of $F_x(37) = \text{pr}(y \leq 37 | x)$ based on the observed data in all the three scenarios. This motivates inference on the additional risk, whose quality is compared with the relevant competitors ANOVA-DDP and F-DMIX in Figure 7.

The results in Figure 7 are in line with those discussed for $n = 500$ in Section 4 of the paper. Specifically, CoMiRe allows accurate learning of $R_A(x, 37)$ both in the correctly specified scenario, and in the cases of model misspecification. The estimation of $R_A(x, 37)$ is particularly precise at low-dose exposures, where quantitative risk assessments typically focus, as can be seen from the bottom panels of Figure 7. As expected, the increased sample size allows further

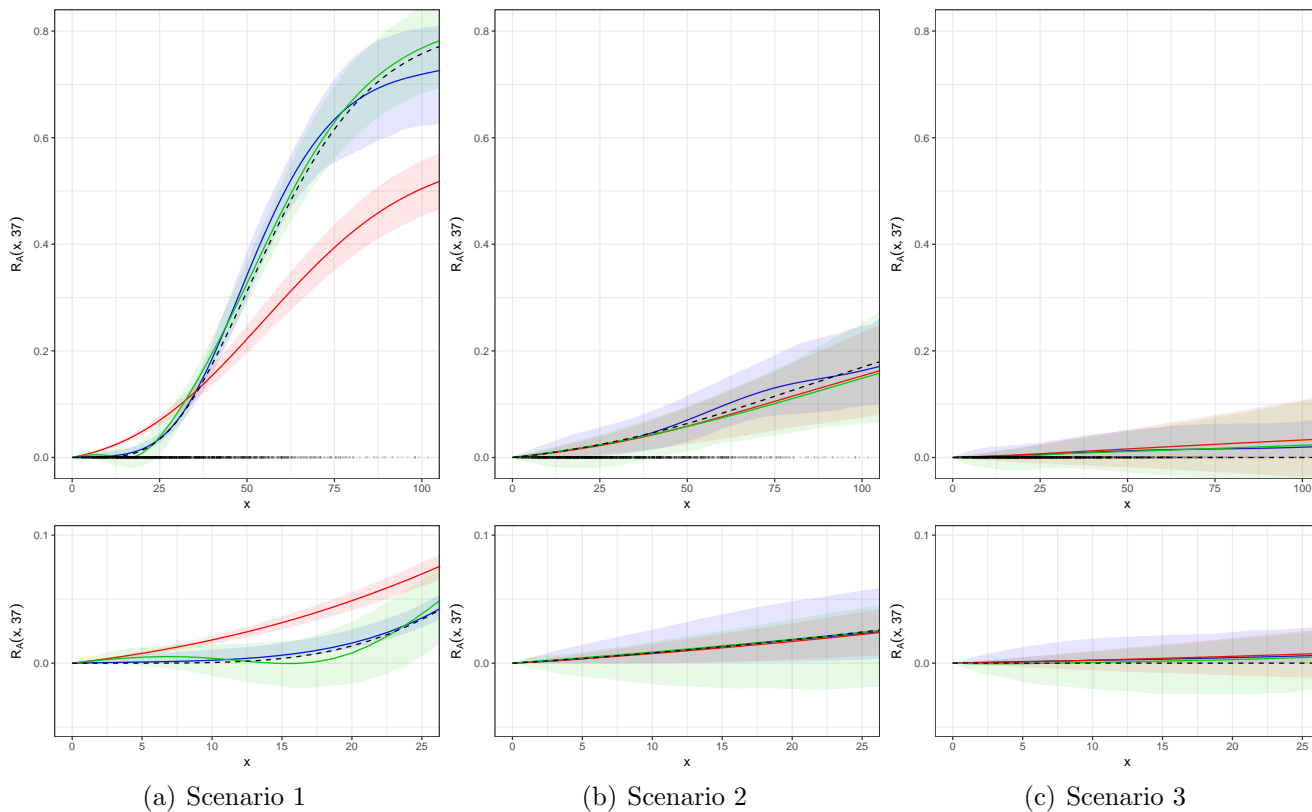


Figure 7: Inference on $R_A(x, 37)$ for $n = 2000$ in the three scenarios. The dashed lines represent the true additional risk function, whereas the red, green, and blue continuous lines denote the posterior mean of the additional risk under ANOVA-DDP, F-DMIX, and CoMiRe, respectively. The shaded areas represent the pointwise 95% posterior credible bands. In the x axis we report the simulated dose exposures. Lower panels provide a zoom on the range of the additional risk typically considered in benchmark dose analysis.

improvements in posterior accuracy and precision for $R_A(x, 37)$ compared to the situation in which $n = 500$.

Code, Data and Tutorial Implementation

Code, data and a tutorial implementation of CoMiRe in the CPP application are available at github.com/tonycanale/CoMiRe.

ACKNOWLEDGEMENT

The authors are grateful to the Editor, the Associate Editor and two referees for the helpful and constructive comments. The first author is supported by the University of Padova under the STARS Grant. This work is also partially supported by grant 1R01-ES028804 of the National Institute of Environmental Health Sciences of the US National Institutes of Health.

REFERENCES

- Calabrese, E. J. and Baldwin, L. A. (2001). U-shaped dose-responses in biology, toxicology, and public health. *Annual Review of Public Health*, 22:15–33.
- Crump, K. S. (1995). Calculation of benchmark doses from continuous data. *Risk Analysis*, 15(1):79–89.
- De Iorio, 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(465):205–215.
- Dunson, D. B. and Park, J.-H. (2008). Kernel stick-breaking processes. *Biometrika*, 95(2):307–323.
- Fronczyk, K. and Kottas, A. (2014). A Bayesian nonparametric modeling framework for developmental toxicity studies. *Journal of the American Statistical Association*, 109(507):873–888.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis (3rd Edition)*. CRC press Boca Raton, FL.
- Good, P. I. (1979). Detection of a treatment effect when not all experimental subjects will respond to treatment. *Biometrics*, 35:483–489.
- He, B., Chen, M., Song, L., and Wang, D. (2010). Mixture normal models in which the proportions of susceptibility are related to dose levels. *Acta Mathematicae Applicatae Sinica, English Series*, 26(3):463–472.

- Hwang, B. S. and Pennell, M. L. (2014). Semiparametric Bayesian joint modeling of a binary and continuous outcome with applications in toxicological risk assessment. *Statistics in Medicine*, 33(7):1162–1175.
- Kodell, R. L. and West, R. W. (1993). Upper confidence limits on excess risk for quantitative responses. *Risk Analysis*, 13(2):177–182.
- Longnecker, M. P., Klebanoff, M. A., Zhou, H., and Brock, J. W. (2001). Association between maternal serum concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies at birth. *The Lancet*, 358(9276):110–114.
- Piegorsch, W. W. and Bailer, A. J. (2005). *Analyzing Environmental Data*. John Wiley & Sons.
- Piegorsch, W. W., West, R. W., Pan, W., and Kodell, R. L. (2005). Low dose risk estimation via simultaneous statistical inferences. *Journal of the Royal Statistical Society: C*, 54(1):245–258.
- Piegorsch, W. W., Xiong, H., Bhattacharya, R. N., and Lin, L. (2014). Benchmark dose analysis via nonparametric regression modeling. *Risk Analysis*, 34(1):135–151.
- Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, 3(4):425–441.
- Razzaghi, M. and Kodell, R. L. (2000). Risk assessment for quantitative responses using a mixture model. *Biometrics*, 56(2):519–527.
- Rigon, T. and Durante, D. (2017). Tractable Bayesian density regression via logit stick-breaking priors. *arXiv:1701.02969*.
- Ritz, C., Baty, F., Streibig, J. C., and Gerhard, D. (2015). Dose-response analysis using R. *PLoS One*, 10(12):e0146021.
- Ritz, C., Gerhard, D., and Hothorn, L. A. (2013). A unified framework for benchmark dose estimation applied to mixed models and model averaging. *Statistics in Biopharmaceutical Research*, 5(1):79–90.
- Rodriguez, A. and Dunson, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis*, 6(1):145–177.

- Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: B*, 73(5):689–710.
- West, R. W. and Kodell, R. L. (1999). A comparison of methods of benchmark-dose estimation for continuous response data. *Risk Analysis*, 19(3):453–459.
- Wheeler, M. W. and Bailer, A. J. (2012). Monotonic Bayesian semiparametric benchmark dose analysis. *Risk Analysis*, 32(7):1207–1218.
- Yu, Z.-F. and Catalano, P. J. (2005). Quantitative risk assessment for multivariate continuous outcomes with application to neurotoxicology: The bivariate case. *Biometrics*, 61(3):757–766.