

# Bayesian Multivariate Isotonic Regression Splines: Applications to Carcinogenicity Studies

Bo Cai and David B. Dunson\*

Biostatistics Branch, National Institute of Environmental Health Sciences,  
P.O. Box 12233, Research Triangle Park, NC 27709, U.S.A.

\*dunson1@niehs.nih.gov

In many applications, interest focuses on assessing the relationship between a predictor and a multivariate outcome variable, and there may be prior knowledge about the shape of the regression curves. For example, regression functions relating dose of a possible risk factor to different adverse outcomes can often be assumed to be nondecreasing. In such cases, interest focuses on (1) assessing evidence of an overall adverse effect; (2) determining which outcomes are most affected; and (3) estimating outcome-specific regression curves. This article proposes a Bayesian approach for addressing this problem, motivated by multi-site tumor data from carcinogenicity experiments. A multivariate smoothing spline model is specified, which accommodates dependency in the multiple curves through a hierarchical Markov random field prior for the basis coefficients, while also allowing for residual correlation. A Gibbs sampler is proposed for posterior computation, and the approach is applied to data on body weight and tumor occurrence.

KEY WORDS: Factor model; Functional data analysis; Monotone curves; Multiple outcomes; Multiplicity problem; Seemingly unrelated regression; Smoothing; Tumor data.

# 1. INTRODUCTION

## 1.1 Background and Motivation

In many applications, data are collected for multiple outcome variables and interest focuses on assessing the functional relationship between a continuous predictor and the mean, adjusting for covariates. In such cases, it is common to have prior knowledge of the direction of the regression functions for one or more of the outcome variables. For example, if the predictor is dose of a potentially adverse exposure, then dose response curves should be non-decreasing for adverse responses, at least after adjusting for important confounding variables. It is well known that incorporating such constraints, when justified, can improve estimation efficiency and power to detect associations (Robertson, Wright and Dykstra, 1988). When outcome data are multivariate, it becomes necessary to not only flexibly model the individual regression functions subject to the non-decreasing constraint but also to characterize dependency.

The problem of flexible nonlinear regression modeling of multiple response data has been the focus of a growing body of literature. Much of the literature has focused on data consisting of repeated observations over time (Brumback and Rice, 1998; Staniswalis and Lee, 1998; Wu and Zhang, 2002, among many others). In the case where data are multivariate normal and the regression functions are expressed as weighted sums of a common set of basis functions, many of the univariate methods generalize directly to the multivariate case (Brown et al., 1998). The literature on seemingly unrelated regressions (SUR) (Zellner, 1962) focuses on the more general case where the basis functions vary for the different outcomes, and the regressions are related through residual correlation (refer to Ng, 2002 and Holmes, Denison, and Mallick, 2002 for recent references). None of these articles considered the incorporation of monotonicity constraints.

For univariate data, a number of approaches have been proposed for frequentist (Mam-

men, 1991; Lee, 1996; Ramsay, 1998) and Bayesian (Lavine and Mockus, 1995; Schmid and Brown, 1999; Holmes and Heard, 2003; Neelon and Dunson, 2004) monotone curve estimation. Our interest is in developing a Bayesian framework for joint estimation and inferences on smooth non-decreasing regression curves for multiple categorical and continuous outcomes. In many applications, interest focuses on not only estimating the regression functions but also assessing evidence of overall increasing trends and the locations of flat regions.

## 1.2 Application: Body Weight and Tumor Occurrence

Our focus is on assessing the relationship between body weight and the occurrence of tumors in different organ sites using data from the National Toxicology Program (NTP) historical control data base. Previous analyses of NTP data have shown evidence of a positive association between body weight at one year of age and the incidence of tumors in certain sites (Seilkop, 1995; Haseman et al., 1997; Parise et al., 2001). Parise et al. (2001) used a semiparametric logistic mixed effects model with a random intercept for study to relate 52 week body weight to the probability of tumor development. They implemented separate analyses for the different tumor sites instead of considering them jointly, and did not consider the incorporation of an order constraint on the curve. Our interest is in jointly modeling the body weight effect on the incidence of the eight most common tumors, including leukaemia, pheochromocytomas, pituitary, mammary, thyroid, subcutaneous, pancreas, and kidney tumors, focusing on male rats for sake of brevity.

Following previous authors in assuming that tumors are unlikely to have progressed to a stage causing morbidity by one year of age (at least in control animals), it is anticipated that tumor incidence is non-decreasing with 52 week body weight, adjusting for animal survival. Our preliminary analysis shown in Figure 1 also confirms the non-decreasing constraint assumption. However, it is not known which tumor types are sensitive to body weight, and an

important issue is to assess how high body weight needs to be before there is an increase in tumor incidence. Figure 1 reports fitted curves and confidence limits from an unconstrained frequentist generalized additive model (GAM) analysis (Hastie and Tibshirani, 1990) using the model:

$$\Pr(y_{ij} = 1 \mid x_i, u_i) = \Phi(\alpha_{1j} + f_j(x_i) + \alpha_{2j}u_i), \quad (1)$$

where  $y_{ij} = 1$  indicates occurrence of a tumor of type  $j$  in rat  $i$ ,  $\Phi(z)$  is the standard normal distribution function,  $\alpha_{1j}$  is an intercept-specific to tumor type  $j$ ,  $u_i = \log t_i/t$  with  $t_i$  age at death for rat  $i$  and  $t = \max\{t_i\}$ ,  $x_i$  is body weight of rat  $i$ , and  $f_j(x)$  is an unknown smooth function. This model was fitted separately for each tumor type using the `gam()` function in S-PLUS. Results are based on data for 1209 male rats from 22 two-year studies.

For each tumor type, the risk of developing a tumor shows an overall increasing trend according to body weight. However, confidence limits are quite wide, even in this initial analysis which did not account for heterogeneity among studies. In addition, many of the curves are flat at low body weights, and it is not clear whether normal weight animals have higher risk than low weight animals. One of our primary goals, in addition to improving precision of estimates through incorporation of non-decreasing constraints on the regression curves, is to identify regions across which increases in body weight have no effect. Such regions are of substantial interest biologically, and may partly address important public health questions, such as whether there is a difference in cancer risk between thin and normal weight or between normal and slightly overweight individuals. In addition, because body weight is an important confounder in studies that assess chemical effects on tumorigenesis, it is important to identify which tumor types are sensitive and how much body weight needs to be modified before there is a change in risk.

### 1.3 Outline of Approach and Related Literature

Motivated by this application, we propose a general Bayesian approach for smooth mono-

tone curve estimation and inferences for multiple outcome data. In particular, we describe a general multivariate isotonic regression factor model for binary outcomes. This approach has some conceptual relationship to the method of Neelon and Dunson (2004), which was developed for univariate isotonic regression. However, motivated by the body weight-tumor incidence application, we propose a different model, prior structure, and computational algorithm, which accommodates multivariate outcome data and allows for study-to-study variability. The model is related to multivariate probit models with a semiparametric isotonic regression structure for the mean and with a hierarchical factor analytic covariance structure. We follow the strategy of Albert and Chib (1993) in introducing latent variables to facilitate posterior computation via a data augmentation Markov chain Monte Carlo (MCMC) algorithm. The factor analytic structure can be chosen motivated by the carcinogenicity application to reduce dimensionality of the covariance matrix relative to the model of Chib and Greenberg (1998).

In Section 2 we discuss the regression model for multiple outcomes and the prior structure for regression coefficients. In Sections 3 and 4, we propose the multivariate isotonic regression model and approach to posterior computation. In Section 5 we describe the multivariate isotonic regression factor model for binary outcomes and the algorithm for posterior computation. In Section 6, we present a simulation study. In Section 7 we apply the approach to the NTP data on body weight and tumor occurrence. In Section 8 we summarize and discuss the results.

## 2. REGRESSION SPLINES FOR MULTIPLE OUTCOMES

Interest focuses on assessing the relationship between a predictor  $x_i$  and a multivariate outcome  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$ , with data collected for subjects  $i = 1, \dots, n$ . We initially assume that  $\mathbf{y}_i$  has a multivariate normal distribution with  $x_i$  the only predictor:

$$\mathbf{y}_i \sim N_p(\boldsymbol{\mu}(x_i), \boldsymbol{\Sigma}), \tag{2}$$

where  $\boldsymbol{\mu}(x) = [f_1(x), \dots, f_p(x)]'$  is a vector of unknown regression functions specific to outcomes  $1, \dots, p$ , respectively,  $\boldsymbol{\Sigma}$  is a covariance matrix, and  $x_i \in \mathcal{X} = [\gamma_0, \gamma_k]$ . Generalizations to allow multiple predictors and binary outcomes will be described in Section 5.

We assume that each curve can be characterized as a linear combination of a common set of potential basis functions so that

$$f_j(x_i) = \beta_{0j} + \sum_{h=1}^k \beta_{hj} B_h(x_i), \quad \text{for } j = 1, \dots, p, \quad (3)$$

where  $\beta_{0j}$  is an intercept for outcome  $j$ , and  $\beta_{hj}$  is a regression coefficient for outcome  $j$  and basis function  $B_h(x)$ . In particular, we will focus on the regression spline basis functions:

$$B_h(x_i) = I(x_i > \gamma_{h-1})[\min(x_i, \gamma_h) - \gamma_{h-1}]^m, \quad (4)$$

where  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)'$  are potential knot points with  $\gamma_0 < \gamma_1 < \dots < \gamma_k$ , and  $m$  gives the order of the spline. We focus on linear splines having  $m = 1$  but the methods generalize directly.

Although we assume a common set of potential basis functions for the different outcomes, the effective number of basis functions can vary since our isotonic regression prior (to be described in Section 3) allows flat regions across which the regression coefficients are constant. The locations of these flat regions can vary for the different outcomes. Markov random field (MRF) formulations originated in spatial statistics and play a central role in the areas such as random graphs, graphical models and statistical physics. To define an MRF, one can specify conditional distributions for each of the elements of a random vector to depend only on neighboring elements. In our case, initially focusing on the unconstrained case, we define an MRF-type prior

$$\pi(\boldsymbol{\beta}_h | \boldsymbol{\beta}_{h-1}, \boldsymbol{\beta}_{h+1}) \stackrel{d}{=} \begin{cases} N_p\left(\left(\frac{\tau}{\tau_0 + \tau}\right)\boldsymbol{\beta}_2, (\tau_0 + \tau)^{-1}\mathbf{I}_p\right) & h = 1 \\ N_p\left(0.5(\boldsymbol{\beta}_{h-1} + \boldsymbol{\beta}_{h+1}), 0.5\tau^{-1}\mathbf{I}_p\right) & h = 2, \dots, k-1 \\ N_p\left(\left(\frac{\tau}{\tau_0 + \tau}\right)\boldsymbol{\beta}_{k-1}, (\tau_0 + \tau)^{-1}\mathbf{I}_p\right) & h = k \end{cases} \quad (5)$$

Under expression (4) with  $m = 1$ , the elements of  $\boldsymbol{\beta}_h = (\beta_{h1}, \dots, \beta_{hp})'$ ,  $h = 1, \dots, k$ , are interpretable as local slopes within interval  $(\gamma_{h-1}, \gamma_h]$  for outcomes  $j = 1, \dots, p$ . The MRF

prior centers  $\beta_h$  on its nearest neighbors,  $\beta_{h-1}$  and  $\beta_{h+1}$ . In the interior bins ( $h = 2, \dots, k - 1$ ), the prior expectation of  $\beta_h$  is simply an average of the slopes in the adjacent bins, with the precision depending on the smoothing parameter  $\tau$ . In the exterior bins ( $h = 1, k$ ), the conditional prior for  $\beta_h$  is proportional to the product of a  $N_p(\beta_h; \mathbf{0}, \tau_0^{-1} \mathbf{I}_p)$  density and a  $N_p$  density centered on the slopes in the adjacent interior bin. This MRF structure allows the interval-specific slopes to be updated by borrowing the information from their neighbors. It also centers the MRF prior for the interval-specific slopes on  $\mathbf{0}$  and smoothes the local slopes across  $\mathcal{X}$  to a degree depending on  $\tau$ . In the limiting case as  $\tau \rightarrow \infty$ ,  $\beta_h$  is constant for  $h = 1, \dots, k$ , and the model reduces to a multiple linear regression model.

One motivation for using this particular MRF structure is that it ensures that the prior variance is the same at both endpoints of  $\mathcal{X}$ . As noted by Denison et al.(2002), the most reasonable default choice in the absence of additional information is to assume equal prior variance at the endpoints. We describe a generalization of this prior to accommodate monotonicity constraints on the regression functions in the next section.

To compute the posterior distribution of  $\beta = (\beta'_1, \dots, \beta'_k)'$  defined above, it is necessary to find the prior expression of  $\beta$ , i.e.  $\pi(\beta)$ . We note that the following expression holds:

$$\pi(\beta) = \pi(\beta_1) \prod_{h=2}^k \pi(\beta_h | \beta_{<h}), \quad (6)$$

where  $\beta_{<h} = \{\beta_j : j < h\}$ . However, the product terms above are not easily obtained. We adopt the recursive algorithm proposed by Bartolucci and Besag (2002) by which the individual terms of (6) can be evaluated as full conditionals of  $\beta_h$  obtained by marginalising over  $\beta_{>h}$ . According to Theorem 1 from Bartolucci and Besag (2002), we can calculate the elements of (6) recursively

$$\pi(\beta_h | \beta_{<h}) = \left[ \int \frac{\pi(\beta_{h+1} | \beta_{<h+1})}{\pi(\beta_h | \beta_{h-1}, \beta_{h+1})} d\beta_{h+1} \right]^{-1}, \quad h = 1, \dots, k - 1 \quad (7)$$

Clearly,  $\pi(\beta_1 | \beta_{<1}) = \pi(\beta_1)$  and  $\pi(\beta_k | \beta_{<k}) = \pi(\beta_k | \beta_{k-1})$ . By substituting the results from

(7) to (6), we obtain the marginal prior of  $\boldsymbol{\beta}$  as follows

$$\pi(\boldsymbol{\beta}) = \begin{cases} \pi(\boldsymbol{\beta}_1) \prod_{t=1}^{s-1} \pi(\boldsymbol{\beta}_{2t} | \boldsymbol{\beta}_{2t-1}, \boldsymbol{\beta}_{2t+1}) \pi(\boldsymbol{\beta}_k | \boldsymbol{\beta}_{k-1}) & k = 2s \\ \pi(\boldsymbol{\beta}_1) \prod_{t=1}^s \pi(\boldsymbol{\beta}_{2t} | \boldsymbol{\beta}_{2t-1}, \boldsymbol{\beta}_{2t+1}) & k = 2s + 1 \end{cases} \quad (8)$$

where  $\pi(\boldsymbol{\beta}_1) = \text{N}(0, \frac{(k-1)\tau_0 + \tau}{[(k-1)\tau_0 + 2\tau]}\mathbf{I})$  and the other terms are from (5). The calculation is straightforward but extensive.

### 3. BAYESIAN ISOTONIC REGRESSION

#### 3.1 Incorporating Non-Decreasing Constraints

Suppose that prior knowledge is available that the regression curves for the first  $p_1$  ( $p_1 \leq p$ ) outcomes are non-decreasing,  $f_j(x_1) \leq f_j(x_2)$  for all  $(x_1, x_2) : x_1, x_2 \in \mathcal{X}$  and  $x_1 < x_2$ , for  $j = 1, \dots, p_1$ . In order to modify the prior of Section 2 to accommodate this constraint, we first replace the regression coefficients  $\boldsymbol{\beta}$  with latent regression coefficients  $\boldsymbol{\beta}^*$  in the MRF prior. We then link  $\boldsymbol{\beta}$  and  $\boldsymbol{\beta}^*$  as follows:

$$\beta_{hj} = \begin{cases} I(\beta_{hj}^* \geq \delta) \beta_{hj}^*, & \text{for } j = 1, \dots, p_1, \\ \beta_{hj} = \beta_{hj}^*, & \text{for } j = p_1 + 1, \dots, p, \end{cases} \quad (9)$$

where  $\delta$  is a small positive constant or threshold below which the slope is effectively zero. A similar link function was used by Neelon and Dunson (2004) for univariate isotonic regression, but they used an AR-1 instead of MRF prior for the latent coefficients. For simplicity, we assume a single threshold  $\delta$  applies to each of the outcomes. In certain cases (e.g., when the scale of the outcomes differs greatly), it may be necessary to include outcome-specific thresholds,  $\delta_j$ . To modify (9) to accommodate non-increasing constraints on  $f_j(x)$ , simply let  $\beta_{hj} = I(\beta_{hj}^* \leq -\delta) \beta_{hj}^*$ .

Under link function (9) and the MRF prior,  $f_j(x)$  is constrained to be non-decreasing for  $j = 1, \dots, p_1$ . In addition, flat regions of  $f_j(x)$ , including the important special case  $H_{0j} : f_j(x) = 0$  for all  $x \in \mathcal{X}$ , are assigned positive probability. This property is useful when interest focuses on assessing evidence that curves are increasing instead of flat. In addition,

by placing mass on the boundary, we obtain a shrinkage estimator of the regression function. Since the result of James and Stein (1961), shrinkage estimates have been shown to have good properties in many cases. In simulation examples, our estimator has exhibited lower risk under squared error loss compared with the corresponding unconstrained estimator for a variety of cases, though uniform domination results are not yet available.

### 3.2 Hierarchical Covariance Structure and Prior Specification

To complete a Bayesian specification of our model, we need to choose prior distributions for the residual covariance  $\Sigma$ , intercept parameters  $\beta_0 = (\beta_{01}, \dots, \beta_{0p})'$ , knots  $\gamma$ , smoothing parameter  $\tau$ , and the threshold  $\delta$ . We assume a priori independence for these different parameter vectors and let

$$\pi(\beta_0) = N_p(\beta_0; \mathbf{0}, s\mathbf{I}_p), \quad \pi(\Sigma) = \mathcal{IW}(\Sigma; v, \mathbf{R}^{-1}), \quad \pi(\tau) = \mathcal{G}(\tau; c, d), \quad \pi(\delta) = \mathcal{G}(\delta; e, f), \quad (10)$$

where  $\mathcal{IW}(\cdot)$  denotes the inverse-Wishart density, and  $s, v, \mathbf{R}, c, d, e, f$  are investigator-specified hyperparameters. With the exception of  $\pi(\delta)$ , these priors are conditionally-conjugate. Issues in choosing these hyperparameters will be discussed in Section 5.

We follow a related approach to Smith and Kohn (1996) and Neelon and Dunson (2004) in choosing a prior for the knots based on selecting from a large number of prespecified potential knot locations (e.g., the unique values of the predictor). In particular, letting  $\gamma$  denote this prespecified vector of knots, our prior for  $\beta$  effectively allows knots to drop out of the model by allowing for flat regions of the regression functions.

## 4. POSTERIOR COMPUTATION

For posterior computation, we propose a hybrid Gibbs-Metropolis algorithm which alternates between the following steps: (1) update  $\tau$  from its full conditional distribution by using rejection sampling; (2) update  $\Sigma$  from its inverse-Wishart full conditional distribution; (3) update the intercept parameters  $\beta_{0j}$  from normal full conditional distributions; and (4)

update  $\delta$  and  $(\beta_{hj}, \beta_{hj}^*)$  in a block by sampling a candidate for  $\delta$  and drawing the  $\beta$ 's from their full conditional distributions.

Let  $\boldsymbol{\theta}_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{kj})'$ ,  $\mathbf{Y}_j = (y_{1j}, y_{2j}, \dots, y_{nj})'$ ,  $\mathbf{y}_{i(-j)} = (y_{i1}, \dots, y_{ij-1}, y_{ij+1}, \dots, y_{ip})'$ , and  $\mathbf{B}_i = [1, B_1(x_i), \dots, B_k(x_i)]'$ . The likelihood can be expressed as the likelihood of  $\mathbf{Y}_j$  conditional on  $\mathbf{Y}_{(-j)}$ , the subvector of  $\mathbf{Y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$  excluding  $\mathbf{Y}_j$ , multiplied by the likelihood of  $\mathbf{Y}_{(-j)}$ . For simplicity in defining the conditional posterior distributions, let

$$\mathbf{Y}_j^* = \mathbf{Y}_j - \left( \begin{bmatrix} \mathbf{y}'_{1(-j)} \\ \vdots \\ \mathbf{y}'_{n(-j)} \end{bmatrix} - \begin{bmatrix} \mathbf{B}'_1 \\ \vdots \\ \mathbf{B}'_n \end{bmatrix} [\boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_{j-1} \boldsymbol{\theta}_{j+1} \cdots \boldsymbol{\theta}_p] \right) \boldsymbol{\Sigma}_{(-j)}^{-1} \boldsymbol{\sigma}_{j(-j)},$$

where  $\boldsymbol{\Sigma}_{(-j)}$  is the submatrix of  $\boldsymbol{\Sigma}$  excluding the  $j$ th row and column, and  $\boldsymbol{\sigma}_{j(-j)}$  is the vector of covariances between  $y_{ij}$  and  $\mathbf{y}_{i(-j)}$ . Then, letting  $\mathbf{B} = [\mathbf{B}_1 \cdots \mathbf{B}_n]'$ , the likelihood of  $\mathbf{Y}_j^*$  is  $\pi(\mathbf{Y}_j^* | \boldsymbol{\theta}_j, \boldsymbol{\Sigma}) = N_n(\mathbf{Y}_j^*; \mathbf{B}\boldsymbol{\theta}_j, \psi_j \mathbf{I}_n)$ , where  $\psi_j = \sigma_{jj} - \boldsymbol{\sigma}'_{j(-j)} \boldsymbol{\Sigma}_{(-j)}^{-1} \boldsymbol{\sigma}_{j(-j)}$ .

The full conditional posterior distribution of  $\beta_{0j}$  is

$$N\left(\beta_{0j}; (s^{-1} + n\psi_j^{-1})^{-1} \mathbf{1}'_n (\mathbf{Y}_j^* - \mathbf{B}_{(-1)} \boldsymbol{\theta}_{j(-1)}), (s^{-1} + n\psi_j^{-1})^{-1}\right), \quad (11)$$

where  $\mathbf{1}_n$  is an  $n \times 1$  vector of ones,  $\mathbf{B}_{(-1)}$  is the submatrix of  $\mathbf{B}$  excluding the 1st column, and  $\boldsymbol{\theta}_{j(-1)}$  is the subvector of  $\boldsymbol{\theta}_j$  excluding the first element.

The full conditional posterior distribution of  $\boldsymbol{\Sigma}$  is

$$\mathcal{IW}\left(\boldsymbol{\Sigma}; v + n/2, (\mathbf{R} + \epsilon' \epsilon)^{-1}\right), \quad (12)$$

where  $\epsilon$  is the  $n \times p$  residual matrix.

The full conditional posterior distribution of  $(\beta_{hj}, \beta_{hj}^*)$  is proportional to

$$I(\beta_{hj} = 0) \hat{P}_{hj} N_{(-\infty, \delta)}(\beta_{hj}^*; E_{hj}, V_{hj}) + I(\beta_{hj} = \beta_{hj}^*) (1 - \hat{P}_{hj}) N_{[\delta, \infty)}(\beta_{hj}^*; \hat{E}_{hj}, \hat{V}_{hj}), \quad (13)$$

where  $E_{hj}$  and  $V_{hj}$  are the conditional prior mean and variance of  $\beta_{hj}^*$ ,  $\mathbf{B}_h$  is the  $(h+1)$ st column vector of  $\mathbf{B}$  (corresponding to basis function  $h$ ),  $\hat{E}_{hj} = \hat{V}_{hj}^{-1} (V_{hj}^{-1} E_{hj} + \psi_j^{-1} \mathbf{B}'_h \mathbf{Y}_{hj}^*)$  and  $\hat{V}_{hj} = (V_{hj}^{-1} + \psi_j^{-1} \mathbf{B}'_h \mathbf{B}_h)^{-1}$  are the unconstrained conditional posterior mean and variance

of  $\beta_{hj}$  derived under  $\beta_{hj} = \beta_{hj}^*$ ,  $\mathbf{Y}_{hj}^* = \mathbf{Y}_j^* - \mathbf{B}\boldsymbol{\theta}_j + \mathbf{B}_h\beta_{hj}^*$ , and the conditional posterior probability of  $\beta_{hj} = 0$  is

$$\hat{P}_{hj} = \left[ 1 + \frac{\Phi(-\delta; \hat{E}_{hj}, \hat{V}_{hj}) N(0; E_{hj}, V_{hj})}{\Phi(\delta; E_{hj}, V_{hj}) N(0; \hat{E}_{hj}, \hat{V}_{hj})} \right]^{-1}.$$

Hence, the full conditional posterior distribution of  $(\beta_{hj}, \beta_{hj}^*)$  is a mixture of two truncated normal distributions, which is straightforward to sample from.

## 5. FACTOR MODELS FOR MULTIPLE BINARY OUTCOMES

### 5.1 Factor Models and Posterior Computation

In our tumor application, we want to assess the multivariate dependence and codependence among different tumors. However, in the model of Section 4, choosing a prior for  $\boldsymbol{\Sigma}$  is challenging, since constraints are needed and the number of unknowns can be large ( $p(p+1)/2$ ). Although the inverse Wishart prior is the default conjugate choice, such a prior is not flexible enough to allow for differential prior knowledge about the different elements of  $\boldsymbol{\Sigma}$  and does not allow one to fix the diagonal elements at one for the binary outcomes for purposes of identifiability. Instead, we use Bayesian factor analysis which has received growing attention in the literature for flexible approaches for parsimonious modeling of covariance structures in biomedical applications involving high dimensional data (eg. gene expression). For an overview of some of the recent issues, refer to the articles of Dunson and Dinse (2002), Dunson (2003), and Lopes and West (2004).

Let  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$  denote the collection of  $p$  binary outcomes for subject  $i = 1, \dots, n$ . Following the approach of Albert and Chib (1993) for the probit model, we define an underlying latent variable vector as  $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})'$ , where  $y_{ij} = I(z_{ij} > 0)$  for  $j = 1, \dots, p$ . Then, an underlying multivariate isotonic regression factor model is of the form

$$\mathbf{z}_i = \boldsymbol{\theta}'\mathbf{B}(\mathbf{x}_i) + \boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i, \quad \text{for } i = 1, \dots, n, \quad (14)$$

where  $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})'$  denotes a predictor vector for subject  $i$ ,  $\mathbf{B}(\mathbf{x}_i) = [1, \mathbf{B}_1(\mathbf{x}_i), \dots, \mathbf{B}_k(\mathbf{x}_i)]'$  denotes the basis function vector for the  $i$ th subject, in which  $\mathbf{B}_h(\mathbf{x}_i) = (B_h(x_{i1}), \dots, B_h(x_{iq}))$  and  $B_h(x)$  is defined in Section 2,  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)$  denotes the regression coefficient matrix, in which  $\boldsymbol{\theta}_j = (\beta_{0j}, \boldsymbol{\beta}'_{1j}, \dots, \boldsymbol{\beta}'_{kj})'$ ,  $\boldsymbol{\beta}_{hj} = (\beta_{hj1}, \dots, \beta_{hjq})'$ ,  $\beta_{lhj}$  is defined in Section 3,  $\boldsymbol{\Lambda}$  is a  $p \times r$  factor loadings matrix,  $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{ir})'$  are independent standard normal factors, and  $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ , with diagonal covariance matrix  $\boldsymbol{\Sigma} = \text{diag}(1, \dots, 1, \sigma_{p_1+1}^2, \dots, \sigma_p^2)$ . The resulting marginal density integrating out  $\boldsymbol{\xi}_i$  is  $\mathbf{z}_i \sim N(\boldsymbol{\theta}'\mathbf{B}(\mathbf{x}_i), \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma})$ .

Let  $\mathbf{Y}$  denote the  $n \times p$  matrix of outcomes, i.e.  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)'$ . The corresponding underlying latent variable matrix is  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$ . Then we can write model (14) as

$$\mathbf{Z} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\Xi}\boldsymbol{\Lambda}' + \boldsymbol{\epsilon}, \quad (15)$$

where  $\mathbf{B} = (\mathbf{B}(\mathbf{x}_1), \dots, \mathbf{B}(\mathbf{x}_n))'$  is the  $n \times (kq + 1)$  basis function matrix,  $\boldsymbol{\theta}$  and  $\boldsymbol{\Lambda}$  are the coefficient matrix and the  $p \times r$  factor loadings matrix defined above, respectively,  $\boldsymbol{\Xi}$  is the  $n \times r$  factor matrix, and  $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)'$  is the  $n \times p$  normal error matrix. Given a random sample of observations  $\mathbf{Y}$  and a prior density  $\pi(\boldsymbol{\Xi}, \boldsymbol{\Lambda}, \boldsymbol{\theta}, \boldsymbol{\Sigma})$  on the parameters of the model, the joint posterior distribution of the parameters and the latent variables  $\mathbf{Z}$  is

$$\begin{aligned} \pi(\mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{\Lambda}, \boldsymbol{\theta}, \boldsymbol{\Sigma} | \mathbf{Y}) &= (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\boldsymbol{\Sigma}^{-1}) \right\} \\ &\times \prod_{i=1}^n \prod_{j=1}^{p_1} \{I(z_{ij} > 0)I(y_{ij} = 1) + I(z_{ij} \leq 0)I(y_{ij} = 0)\} \\ &\times \pi(\boldsymbol{\Xi}, \boldsymbol{\Lambda}, \boldsymbol{\theta}, \boldsymbol{\Sigma}), \end{aligned} \quad (16)$$

where  $\text{tr}(A)$  is the trace of matrix  $A$ .

In order for the model to be identified, we need to place restrictions on the factor loadings matrix  $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_p)'$ . We follow the common convention of using a lower triangular structure with  $\lambda_{jk} = 0$  for all  $k > j$ ,  $\lambda_{kk} > 0$  for  $k = 1, \dots, r$ , and  $\lambda_{jk} \in \Re$  for  $j > k$ . Assuming a priori independence, we then choose  $\pi(\lambda_{kk}) \propto N(E_{0,kk}, V_{0,kk})I(\lambda_{kk} > 0)$  for  $k = 1, \dots, r$ , and  $\pi(\lambda_{jk}) = N(E_{0,jk}, V_{0,jk})$  for  $j = k + 1, \dots, p$  and  $k = 1, \dots, r$ . For

$\sigma_j^2, j = p_1 + 1, \dots, p$ , we choose inverse gamma priors,  $\pi(\sigma_j^2) = \mathcal{IG}(c_1, d_1)$ . The priors for the other parameters are chosen as in Section 3.2. The full conditional posterior distributions for all parameters and latent variables can be derived from (16) (refer to the Appendix).

The MCMC algorithm of Section 4 can be modified by replacing step (2) with steps for updating the underlying normal variables  $\{z_{ij}\}$ , latent factors  $\{\xi_i\}$ , and factor loadings from their conjugate full conditional posterior distributions, while also modifying the conditional distributions used in steps (1), (3) and (4) slightly. Samples from the joint posterior distribution of the parameters and the latent variables are generated by repeating these steps for a large number of iterations after apparent convergence.

It is of interest to assess how high body weight needs to be before there is an increase in tumor incidence relative to low weight animals. For the  $j$ th tumor type, this body weight corresponds to the first increase in  $f_j(\cdot)$ , which corresponds to the weight  $b_j = \min\{x_i : x_i \in (\gamma_{h-1}, \gamma_h], \beta_{hj} > 0\}$ . The posterior densities of  $b_j$ , for  $j = 1, \dots, p$ , can be estimated directly from the MCMC output. Also, of substantial interest is the posterior probability that the  $j$ th tumor type exhibits an overall increase in incidence across the range of body weights observed in the data base. The null hypothesis of no increase in incidence for the  $j$ th tumor can be expressed as:  $H_{0j} : \beta_{1j} = \dots = \beta_{kj} = 0$ , and we can easily estimate  $\Pr(H_{0j} | \text{data})$  by averaging model indicators across MCMC iterates collected after burn-in. However, it is not enough to assess whether there is an overall increasing trend, we want to assess whether incidence increases only for very heavy animals in the upper range of the distribution or there is an increase in risk even for average weight animals. This question has design implications, since the very heavy animals can potentially be removed by selective breeding (there has been a trend towards fatter animals provided by the animal breeders). In addition, to the extent that such trends reflect human health, it is certainly interesting whether very thin individuals are at lower risk than average weight animals. Such questions can be addressed using our Bayesian approach by estimating posterior probabilities of an

increase within particular ranges of body weight, which is straightforward by again averaging appropriately-defined model indicators.

## 5.2 Generalizations: Covariates and Multilevel Models

Let  $\mathbf{w}_i = (w_{i1}, \dots, w_{iC})'$  denote a  $C \times 1$  vector of additional covariates for subject  $i$ , which are assumed to have linear effects. The model (15) can be extended as

$$\mathbf{Z} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\Xi}\boldsymbol{\Lambda}' + \mathbf{W}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (17)$$

where  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)'$  and  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p)$ , with  $\boldsymbol{\alpha}_j = (\alpha_{1j}, \dots, \alpha_{Cj})'$ . Assume that the prior for  $\boldsymbol{\alpha}_j$  is  $N_p(\mathbf{0}, c_0 I_p)$ . The full conditional posterior distribution for  $\alpha_{cj}$ ,  $c = 1, \dots, C$ ,  $j = 1, \dots, p$  is

$$\pi(\alpha_{cj} | \mathbf{Z}_j, \boldsymbol{\Xi}, \boldsymbol{\Lambda}_j, \boldsymbol{\theta}_j, \mathbf{W}_{(-c)}, \boldsymbol{\alpha}_{(-c)j}, c_0, \sigma_j^2) \propto N(C^* \mathbf{W}'_c (\mathbf{Z}_j - \mathbf{B}\boldsymbol{\theta}_j - \boldsymbol{\Xi}\boldsymbol{\Lambda}'_j - \mathbf{W}_{(-c)}\boldsymbol{\alpha}_{(-c)j}), C^*),$$

where  $\mathbf{W}_c$  denotes the  $c$ th column of  $\mathbf{W}$ ,  $\mathbf{W}_{(-c)}$  denotes  $\mathbf{W}$  without the  $c$ th column,  $\boldsymbol{\alpha}_{(-c)j}$  denotes  $\boldsymbol{\alpha}_j$  without the  $c$ th element, and  $C^* = (c_0^{-1} + \mathbf{W}'_c \mathbf{W}_c \sigma_j^{-2})^{-1}$ .

In the tumor application, we have data from multiple studies and it is necessary to accommodate possible heterogeneity. For this reason, we generalize our model to include study-specific random effects for the different outcome types. In particular, we include the random effects  $\boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{ip})'$  ( $i = 1, \dots, n$ ), where  $i$  now indexes the study (or cluster),  $j$  indexes subjects within clusters ( $j = 1, \dots, n_i$ ), and  $\zeta_{iu} \stackrel{ind}{\sim} N(0, \sigma_{2u}^2)$  for outcomes  $u = 1, \dots, p$ . The data are of the form  $(\mathbf{y}_{ij}, \mathbf{x}_{ij})$  ( $i = 1, \dots, n; j = 1, \dots, n_i$ ) where  $n_i$  denotes the number of observations in cluster  $i$ . Then we can rewrite the multivariate isotonic regression factor model (14) as follows

$$\mathbf{z}_{ij} = \boldsymbol{\theta}'\mathbf{B}(\mathbf{x}_{ij}) + \boldsymbol{\Lambda}\boldsymbol{\xi}_{ij} + \boldsymbol{\zeta}_i + \boldsymbol{\epsilon}_{ij}. \quad (18)$$

In cluster  $i$ , the model (18) can also be written as

$$\mathbf{z}_i = \mathbf{B}(\mathbf{x}_i)\boldsymbol{\theta} + \boldsymbol{\Xi}_i\boldsymbol{\Lambda} + \mathbf{1}_{n_i}\boldsymbol{\zeta}'_i + \boldsymbol{\epsilon}_i, \quad (19)$$

where  $\mathbf{z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{in_i})'$ ,  $\Xi_i$  is the  $n_i \times r$  factor matrix, and  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})'$  is the  $n_i \times p$  normal error matrix. Assuming that the prior for  $\sigma_{2u}^{-2}$  is  $\mathcal{G}(\kappa_1, \kappa_2)$ , we can obtain the full conditional for  $\sigma_{2u}^{-2}$  as follows

$$\pi(\sigma_{2u}^{-2} | \zeta) \propto \mathcal{G}\left(\kappa_1 + \frac{n}{2}, \kappa_2 + \frac{1}{2} \sum_{i=1}^n \zeta_{iu}^2\right).$$

The full conditional for  $\zeta_i$  is

$$\pi(\zeta_i | \mathbf{z}_i, \xi_i, \Lambda, \theta) \propto \text{N}\left(V_{\zeta_i} \Sigma_1^{-1} (\mathbf{z}_i - \mathbf{B}(\mathbf{x}_i) \theta - \Xi_i \Lambda)' \mathbf{1}_{n_i}, V_{\zeta_i}\right),$$

where  $V_{\zeta_i} = (\Sigma_2^{-1} + n_i \Sigma_1^{-1})^{-1}$ ,  $\Sigma_1$  and  $\Sigma_2$  are covariance matrices of  $\epsilon_{ij}$  and  $\zeta_i$ , respectively.

These full conditional distributions are easily simulated.

## 6. SIMULATION STUDY

We conducted a simulation study to evaluate the behavior of the proposed procedure. We first compare the estimates of our proposed approach without the SUR component with the Ramsay (1998) smooth monotone function estimator. We generated data from  $y_i \sim \text{N}(f(x_i), 1)$  with  $f(x)$  chosen as 1) flat curve, 2) linear curve, 3) step curve, and 4) cubic curve, with the predictor values generated from  $x_i \sim \text{Uniform}(0, 10)$ , for  $i = 1, \dots, n = 200$ . We chose priors for intercept  $\beta_0$  and the error variance as  $\text{N}(0, 10)$  and  $\mathcal{IG}(0.1, 0.1)$ , respectively. The priors for  $\tau$  and  $\delta$  are chosen as  $\mathcal{G}(2, 1)$  and  $\mathcal{G}(1.25, 25)$ , respectively. We also set grid number  $k$  to be 100.

We ran our MCMC algorithm for 10,000 iterations after a 5,000 burn-in. The samples passed convergence diagnostic tests in CODA (Best et al., 1995). We chose a set of MCMC samples for inference by thinning every 5 steps in the chain. We also conduct a sensitivity analysis to the choice of prior by rerunning all of the analyses for different choices of the hyperparameters, including (i) priors with variance  $\times 2$ ; (ii) priors with variance  $/2$ ; and (iii) priors with different means in a range consistent with our prior expectation.

Figure 2 presents the estimates for the four cases above. For the flat curve, the posterior mean from the proposed approach was almost flat and closer to the true curve than the Ramsay estimator which had a slight increasing trend in the low range. For the linear curve, the estimates of our proposed approach and the Ramsay estimator was basically the same which were close to the true curve, though they are not linear. For the step curve, it is clear that our Bayesian estimator successfully captured the first jump of the true curve but smoothed out the second jump. In contrast, the Ramsay estimator failed to capture the jumps. For the cubic curve, our estimate was closer to the true curve than the Ramsay estimate, which was slightly over-smoothed.

We also studied trivariate outcome data generated from a multivariate normal distribution  $N(f(x), \Sigma)$  with  $f(x)$  for all outcomes chosen to be (1) flat curves; (2) linear curves with different slopes; (3) step curves with a variety of different shapes; and (4) non-linear curves, and for simplicity, with common covariance matrix

$$\Sigma = \begin{pmatrix} 4 & 1.5 & 3.6 \\ 1.5 & 4 & 0.8 \\ 3.6 & 0.8 & 4 \end{pmatrix}.$$

We assigned independent  $N(1, 100)$  priors truncated below by 0 for identifiability for the factor loadings,  $\mathbf{\Lambda} = (\lambda_1, \lambda_2, \lambda_3)'$ . Priors for the remaining parameters and computational details are similar to the settings in the first simulation which are omitted due to space.

Figure 3 shows the estimates from the linear regression model (LM), generalized additive model (GAM) with identity link, Ramsay (1998) smooth monotone spline method, and our proposed method with SUR and without SUR, in which data were fitted for each outcome separately in GAM and Ramsay method. In case 1 and case 2, linear models did the best as expected, and our approach did a good job as well which is much better than GAM and Ramsay method. In case 3, our method successfully caught the right shape of the true curves while GAM and Ramsay failed. In case 4, our approach fitted the true curves well, which is slightly better than GAM and Ramsay method. Not surprisingly, the linear model failed

in case 3 and 4. We also note that separate fit from our approach without SUR has less accuracy due to the neglect of dependence among outcomes.

Figure 4 presents the posterior densities for  $\tau$  based on 20 different values of the first hyperparameter of  $(c, d)$  ranged from 1 to 4, and the corresponding curves to these hyperparameters. The plot shows the robustness of the curves to this prior specification.

## 7. MULTIPLE TUMOR SITE DATA APPLICATION

The data used in our application are from the National Toxicology Program (NTP) historical control database for long-term rodent carcinogenicity bioassays. In this article we focus on 1209 male rats, with data consisting of individual animal body weights in grams at 12 months, age at death, and 0/1 tumor indicators for 8 of the most frequently occurring tumor types. The range of body weights for male rats is between 287.9 and 581.7 grams, with a median of 479.7. We incorporate age at death as a covariate into the model (Dinse and Lagakos, 1983). The objective of the study is to model the 8 binary outcomes as a function of body weight, allowing for correlation among the outcomes.

Our strong a priori belief is that tumor incidence is non-decreasing as 12 month body weight increases, though certain tumor types may not depend on body weight and we are uncertain how high body weight needs to be before there is an increase in incidence. We analyze the data using our Bayesian approach in which a non-decreasing constraint is incorporated in the regression function for each outcome. We consider a two level Bayesian multivariate isotonic regression model, with a single animal-specific factor,  $\xi_i$ , measuring the susceptibility of animal  $i$  to tumor development and with study-specific random effects,  $\zeta_i = (\zeta_{i1}, \dots, \zeta_{i8})$ , for each tumor type. Since the outcomes are binary, the  $\epsilon_i$  in (14) becomes an independent normal random variate vector with mean zero and identity matrix  $\Sigma$ , i.e.  $\Sigma = \text{diag}(1, \dots, 1)$ .

To specify the degree of smoothing in the MRF prior, we chose  $\pi(\tau) = \mathcal{G}(2, 1)$  and

$\tau_0 = 2$ , which corresponds to a moderate to high degree of smoothing. The factor loadings,  $\mathbf{\Lambda} = (\lambda_1, \dots, \lambda_8)'$ , are assigned independent  $N(1, 100)$  priors truncated below by 0 for identifiability. Variance of 100 corresponds to a highly diffuse specification for these parameters. For the prior on the parameter  $\delta$ , which controls the probabilities allocated to flat regions of the regression functions, we chose  $\mathcal{G}(1.5, 30)$ . This prior is centered on  $E(\delta) = 0.05$  with standard deviation 0.04. A value of 0.05 or less for the slope was considered to be effectively 0 in the context of the application (note that the body weights were standardized prior to analysis). The prior for  $\beta_0$  is chosen as  $N(\mathbf{0}, 10\mathbf{I})$  with 10 to correspond to a vague specification for the intercept. A diffuse prior for the survival effect coefficients,  $\alpha$ , is chosen as  $N(\mathbf{0}, 10\mathbf{I})$ . Finally, the prior for  $\sigma_{2u}^{-2}$  is chosen as  $\mathcal{G}(0.5, 50)$ . We chose the potential knots  $\gamma$  to correspond to an equally-spaced grid of  $k = 100$  points spanning the range of body weights observed in the data base.

We ran our MCMC algorithm for 50,000 iterations after a 10,000 burn-in. Diagnostic tests showed rapid convergence and efficient mixing. A sample of size 2,000 was obtained by thinning the MCMC chain by a factor of 25. To conduct a sensitivity analysis to the choice of prior, we reran all of the analyses for different choices of the hyperparameters, including (i) priors with variance  $\times 2$ ; (ii) priors with variance  $/2$ ; and (iii) priors with different means in a range consistent with our prior expectation. Although we do not report the results in detail, inferences tended to be quite robust to the prior specification. Table 1 shows the probabilities of getting the tumor for an animal with the mean body weight, where the hyperparameters  $(c, d)$  for the prior of  $\tau$  are fixed at  $(2, 1)$  while the hyperparameters  $(e, f)$  for the prior of  $\delta$  are chosen with different values. It appears that with the mean body weight, an animal has a big chance to have leukaemia and kidney tumors.

To compare our Bayesian results with a simple frequentist regression analysis, we also fit the data using an unconstrained probit GAM separately for each tumor type without accounting for study-to-study heterogeneity. Figure 5 shows the comparison of the estimated

tumor type-specific incidence curves as a function of body weight for our Bayesian model and the GAM model, together with the pointwise 95% credible intervals from our Bayesian analysis. The estimated posterior mean of the tumor incidence curve for each tumor site under our Bayesian approach is similar to that from the unconstrained frequentist GAM analysis, which is reassuring in that it suggests that the order restriction and prior structure is not introducing a high degree of systematic estimation bias. A primary difference (in addition to facilitating inferences on increasing regions of the regression curves) is that the Bayesian isotonic regression approach yields much more precise estimates compared with the unconstrained GAM model. In particular, the 95% credible intervals are much narrower than those from the GAM model, even accounting for study-to-study variability (which tends to increase uncertainty and widen the intervals). This is not surprising, since large gains in efficiency due to incorporation of a non-decreasing constraint are expected.

The survival-adjusted probability of getting a tumor of a particular type tends to increase non-linearly across the range of body weights observed in the data base, with the shape of the regression function varying somewhat for the different tumor types under consideration. Our Bayesian estimates maintain a strict non-increasing shape, while the frequentist unconstrained estimates exhibit small downturns at intermediate levels of body weight for thyroid tumor, subcutaneous tumor and pancreas tumor. These deviations are all small and are likely to reflect estimation uncertainty instead of evidence against the order restriction. The frequentist GAM analysis gives non-significant p-values for trend tests for all tumor types while our Bayesian approach provides clear evidence of increasing response trends, with the posterior probability of the null hypotheses being less than 0.03 for each tumor type. The increasing trend in tumor incidence with increasing body weight for mammary tumor and pituitary tumor is consistent with the results of Seilkop (1995) and Haseman et al. (1997). We have also detected significant body weight-tumor incidence associations for additional tumors not flagged in these earlier analyses. In particular, pheochromocytoma tumor, thy-

roid tumor and pancreas tumor, whose posterior probability of the null hypotheses is 0.015, 0.023 and 0.026, respectively, were not judged as body weight dependent by previous authors. This is likely due to increased power of the incorporation of the non-decreasing constraint.

A clear advantage of our analysis is that we have a formal framework for assessing evidence of trends, not only overall across the range of observed body weights but also within particular regions. Figure 6 presents the estimated probability of an increase in the tumor-specific incidence by each body weight relative to animals of low body weight. Formally, these estimated curves are simply the posterior probabilities that  $f_j(x)$  exhibits at least one increase (i.e., is not flat) within the interval  $[\gamma_0, x]$ . These probability curves can be calculated easily from the MCMC output. Increases in body weight up to 425 grams appear to increase kidney tumor incidence significantly, but the curve flattens out at higher body weights at which essentially all the rats have tumors. The high incidence of kidney tumors in male rats is a well known phenomenon, which leads to substantial mortality. For thyroid and pancreas tumors, probabilities of an increase are quite stable at the range of body weight between 430 and 500 grams, implying that these two tumors do not increase in incidence when body weight increases within this range. In contrast, body weight appears not to affect the increase of subcutaneous tumors at the body weight range of 470 and 525 grams. We can also obtain some detailed information in Figure 6. At the lower range of body weight of being below 425 grams, the increase of leukaemia, thyroid tumor and kidney tumor is more sensitive to body weight than that of the other tumors. In contrast, the increase of subcutaneous tumor and pancreas tumor shows a stronger link to the higher body-weight range over 525 grams than that of the others. These results are robust to the prior specification.

To allow for heterogeneity among studies in the historical control database, we have incorporated study-specific random effects for each tumor. It is interesting to assess the level of heterogeneity among studies in the different tumor types, adjusting for differences

in animal survival and body weight. NTP investigators are worried about a trend towards increased tumor incidence over time attributable to increases in body weight, and it is not clear how much residual study-to-study variability remains after taking into account differences in animal weight. This data set is based on studies conducted before NTP changed to a new diet in an attempt to limit heterogeneity among studies. Based on examining posterior distributions, most of the random effect variances are close to 0.10, which represents a small to moderate degree of heterogeneity on the scale of the linear predictor (although this is not a formal test).

In addition, by using our Bayesian approach, we can estimate the threshold body weight corresponding to the lowest weight such that there is an increased incidence relative to animals at the bottom of the weight range. Table 2 presents posterior summaries of the threshold body weight for each tumor. These threshold body weights vary substantially for the different tumor types, which is consistent with previous studies showing that the body weight effect is strongly dependent on tumor type. Interestingly, leukaemia seems to be particularly sensitive, with the estimated increase in incidence occurring starting at a body weight of 321.75 grams, which is only in the 0.08th percentile of the distribution. This is consistent with the estimates presented in Figure 1. Hence, for certain tumors of major public health concern, it appears that even normal weight animals are of increased risk relative to thin animals. A verification of the generalizability of this result to humans through epidemiologic studies would be interesting. Kidney tumors also exhibit an increase starting at low body weights, but it is less clear that this result would pertain to humans, given that almost all male rats develop kidney tumors.

## 8. DISCUSSION

In this article we have proposed a Bayesian approach for multi-site tumor data from carcinogenicity experiments. A multivariate smoothing spline model is specified, which ac-

commodates dependency in the multiple curves through a hierarchical Markov random field prior for the basis coefficients, while also allowing for residual correlation through a multi-level factor model. We have described a generalized multivariate isotonic regression factor model for binary outcomes. By incorporating positive prior probability to flat regions instead of using a strictly increasing function, we allow inferences on the occurrence and local of regions across which there is no effect of a predictor. The resulting regression curve estimator has a shrinkage-type structure, which may limit problems with over-estimation bias observed for certain restricted estimators.

This approach was motivated by the NTP application to body weight and tumor incidence described in Section 6. Some of the appealing features of our approach were illustrated in this application. First, precision of estimates can be improved substantially through incorporation of non-decreasing constraints on the regression curves. Second, our approach allows one to investigate the risk of an increase in incidence attributable to increasing body weight by a given amount, which was not fully addressed by previous studies. Such inferences are of primary interest in the tumor application, since it is important to assess whether or not body weight has an effect not just for obese animals but also across the normal range of body weights observed. Based on our results, it is clear that even normal weight animals can have increased risk of certain tumors, most notably leukaemia but for other tumor types as well. Potentially, the same type of methods can be applied to epidemiologic data collected in studies verifying the occurrence of such effects in humans.

Although the methods presented in this article are focused heavily on the analysis and interpretation of the data from rodent carcinogenicity experiments, they can be applied in a wide variety of settings. It is also straightforward to extend our method to deal with mixed categorical and continuous outcomes. In most epidemiologic and toxicologic studies involving a potentially-adverse exposure, such as dose of a chemical, it is reasonable to assume a non-decreasing regression curve after adjusting for important confounding factors, such as age.

In such settings, in which one collects multivariate mixed discrete and continuous outcome data, it is straightforward to apply the approach described in this article to perform inference on the unknown regression function.

## APPENDIX

The full conditional posteriors for the parameters in Section 5.2:

- The full conditional posterior distribution for  $\boldsymbol{\xi}_i$  is  $N(V_{\boldsymbol{\xi}_i} \boldsymbol{\Lambda}' \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\theta}' \mathbf{B}(\mathbf{x}_i)), V_{\boldsymbol{\xi}_i})$ , where  $V_{\boldsymbol{\xi}_i} = (\mathbf{I}_r + \boldsymbol{\Lambda}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda})^{-1}$ .
- The full conditional for  $\sigma_j^2$  is  $\mathcal{IG}(c_1 + \frac{n}{2}, d_1 + \frac{1}{2}(\mathbf{Z}_j - \mathbf{B}\boldsymbol{\theta}_j - \boldsymbol{\Xi}\boldsymbol{\Lambda}'_j)'(\mathbf{Z}_j - \mathbf{B}\boldsymbol{\theta}_j - \boldsymbol{\Xi}\boldsymbol{\Lambda}'_j))$ , for  $j = p_1 + 1, \dots, p$ , where  $\mathbf{Z}_j$  denotes the  $j$ th column of  $\mathbf{Z}$  and  $\boldsymbol{\theta}_j$  denotes the  $j$ th column of  $\boldsymbol{\theta}$ .
- The full conditional for  $\lambda_{kk}$  is  $N(V_{kk}(V_{0,kk}^{-1} E_{0,kk} + \sigma_k^{-2} \boldsymbol{\Xi}'_k (\mathbf{Z}_k - \mathbf{B}\boldsymbol{\theta}_k)), V_{kk}) I(\lambda_{kk} > 0)$ , for  $k = 1, \dots, r$ , where  $V_{kk} = (V_{0,kk}^{-1} + \sigma_k^{-2} \boldsymbol{\Xi}'_k \boldsymbol{\Xi}_k)^{-1}$  and  $\boldsymbol{\Xi}_k$  denotes the  $k$ th column of  $\boldsymbol{\Xi}$ . The full conditional distribution for  $\lambda_{jk}$ ,  $j = k + 1, \dots, p$ , is  $N(V_{jk}(V_{0,jk}^{-1} E_{0,jk} + \sigma_j^{-2} \boldsymbol{\Xi}'_k (\mathbf{Z}_j - \mathbf{B}\boldsymbol{\theta}_j)), V_{jk})$ , where  $V_{jk} = (V_{0,jk}^{-1} + \sigma_j^{-2} \boldsymbol{\Xi}'_k \boldsymbol{\Xi}_k)^{-1}$ .
- The full conditional for  $\tau$  is:

$$\begin{cases} \mathcal{G}(\tau | c_1 + (s-1)pq, d_1 + \sum_{t=1}^{s-1} \text{tr}[\mathbf{C}'\mathbf{C}]) \pi(\boldsymbol{\beta}_1^*) \pi(\boldsymbol{\beta}_k^* | \boldsymbol{\beta}_{k-1}^*) & k = 2s \\ \mathcal{G}(\tau | c_1 + spq, d_1 + \sum_{t=1}^s \text{tr}[\mathbf{C}'\mathbf{C}]) \pi(\boldsymbol{\beta}_1^*) & k = 2s + 1 \end{cases}$$

where  $\mathbf{C} = (\boldsymbol{\beta}_{2t}^* - \frac{1}{2}(\boldsymbol{\beta}_{2t-1}^* + \boldsymbol{\beta}_{2t+1}^*))$ .

- The full conditional for  $\beta_{0j}$  is  $N(S \mathbf{1}_n (\mathbf{Z}_j - B_{(-1)} \boldsymbol{\theta}_{j(-1)} - \boldsymbol{\Xi} \boldsymbol{\Lambda}'_j), S)$ , where  $S = (s^{-1} + n\sigma_j^{-2})^{-1}$ ,  $\mathbf{B}_{(-1)}$  denotes the basis function matrix without the first column and  $\boldsymbol{\theta}_{j(-1)}$  denotes the  $\boldsymbol{\theta}_j$  without the first element.
- The full conditional for  $(\beta_{hjl}, \beta_{hjl}^*)$  is  $I(\beta_{hjl} = 0) \hat{P}_{hjl} N_{(-\infty, \delta)}(\beta_{hjl}^* | E_{hjl}, V_{hjl}) + I(\beta_{hjl} = \beta_{hjl}^*) (1 - \hat{P}_{hjl}) N_{[\delta, \infty)}(\beta_{hjl}^* | \hat{E}_{hjl}, \hat{V}_{hjl})$ , where (1)  $E_{hjl}$  and  $V_{hjl}$  are the conditional prior

mean and variance of  $\beta_{hjl}^*$  derived conditional on  $\beta_{hjl} = 0$ ; (2)  $\hat{E}_{hj} = \hat{V}_{hjl}(V_{hjl}^{-1}E_{hjl} + \sigma_j^{-2}\mathbf{B}'_m\mathbf{Z}_{j(-m)}^*)$  and  $\hat{V}_{hjl} = (V_{hjl}^{-1} + \sigma_j^{-2}\mathbf{B}'_m\mathbf{B}_m)^{-1}$  are the unconstrained conditional posterior mean and variance of  $\beta_{hjl}$  derived under  $\beta_{hjl} = \beta_{hjl}^*$ , in which  $\mathbf{B}_m$  is the  $m$ th column vector of  $\mathbf{B}$ ,  $m = (h-1)q + l$ ,  $\mathbf{B}_{-m}$  is the submatrix of  $\mathbf{B}$  excluding the  $m$ th column,  $\boldsymbol{\theta}_{j(-m)}$  is the subvector of  $\boldsymbol{\theta}_j$  excluding the  $m$ th element, and  $\mathbf{Z}_{j(-m)}^* = \mathbf{Z}_j - \mathbf{B}_{-m}\boldsymbol{\theta}_{j(-m)} - \boldsymbol{\Xi}\boldsymbol{\Lambda}'_j$ ; (3)  $\hat{P}_{hj}$  is the conditional posterior probability of  $\beta_{hjl} = 0$ .

- The full conditional for underlying latent variable  $z_{ij}$ , for  $j = 1, \dots, p_1$ , is  $N(\mathbf{B}(\mathbf{x}_i)\boldsymbol{\theta}_j + \boldsymbol{\Lambda}_j\boldsymbol{\xi}_i, 1)[I(z_{ij} > 0)I(y_{ij} = 1) + I(z_{ij} \leq 0)I(y_{ij} = 0)]$ .

## REFERENCES

- Albert, J.H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669-679.
- Bartolucci, F. and Besag, J. (2002), "A Recursive Algorithm for Markov Random Fields," *Biometrika*, 89, 724-730.
- Best, N.G., Cowles, M.K. and Vines, S.K. (1995), *CODA manual version 0.30*. MRC Biostatistics Unit, Cambridge, UK.
- Brown, P.J., Vannucci, M., and Fearn, T. (1998), "Multivariate Bayesian Variable Selection and Prediction," *Journal of the Royal Statistical Society, Series B*, 60, 627-641.
- Brumback, B.A., and Rice, J.A. (1998), "Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves," *Journal of the American Statistical Association*, 93, 961-976.
- Chib, S. and Greenberg, E. (1998), "Analysis of Multivariate Probit Models," *Biometrika*, 85, 347-361.

- Denison, D., Holmes, C., Mallick, B. and Smith, A.F.M. (2002), *Bayesian Methods for Nonlinear Classification and Regression*. London: Wiley.
- Dinse, G.E., and Lagakos, S.W. (1983), "Regression Analysis of Tumor Prevalence Data," *Applied Statistics*, 32, 236-248.
- Dunson, D.B., and Dinse, G.E. (2002), "Bayesian Models for Multivariate Current Status Data with Informative Censoring," *Biometrics*, 58, 79-88.
- Dunson, D.B. (2003), "Dynamic Latent Trait Models for Multidimensional Longitudinal Data," *Journal of the American Statistical Association*, 98, 555-563(9).
- Haseman, J.K., Young, E., Eustis, S.L., and Hailey, J.R. (1997), "Body Weight-Tumor Incidence Correlations in Long-Term Rodent Carcinogenicity Studies," *Toxicologic Pathology*, 25, 256-263.
- Hastie, T.J. and Tibshirani, R.J. (1990), *Generalized Additive Models*. London: Chapman & Hall.
- Holmes, C.C., Denison, D.G.T, and Mallick, B.K. (2002), Accounting for Model Uncertainty in Seemingly Unrelated Regressions," *Journal of Computational and Graphical Statistics*, 11, 533-551.
- Holmes, C.C. and Heard, N.A. (2003), "Generalised Monotonic Regression using Random Change Points," *Statistics in Medicine*, 22, 623-638.
- James, W., and Stein, C.M. (1961), "Estimation with Quadratic Loss," *Proceedings of the 4th Berkeley Symposium 1*, 361-379.
- Lavine, M. and Mockus, A. (1995), "A Nonparametric Bayes Method for Isotonic Regression," *Journal of Statistical Planning and Inference*, 46, 235-248.

- Lee, C.I.C. (1996), "On Estimation for Monotone Dose-Response Curves," *Journal of the American Statistical Association*, 91, 1110-1119.
- Lopes, H.F. and West, M. (2004), "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14, 41-67.
- Mammen, E. (1991), "Nonparametric Regression under Qualitative Smoothness Assumptions," *Annals of Statistics*, 19, 741-759.
- Neelon, B. and Dunson, D.B. (2004), "Bayesian Isotonic Regression and Trend Analysis," *Biometrics*, 60, 398-406.
- Ng, V.M. (2002), "Robust Bayesian Inference for Seemingly Unrelated Regressions with Elliptical Errors," *Journal of Multivariate Analysis*, 83, 409-414.
- Parise, H., Wand, M.P., Ruppert, D., and Ryan, L.M. (2001), "Incorporation of Historical Controls Using Semiparametric Mixed Models," *Applied Statistics*, 50, 31-42.
- Ramsay, J.O. (1998), "Estimating Smooth Monotone Functions," *Journal of the Royal Statistical Society, Series B*, 60, 365-375.
- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*. Wiley, New York.
- Schmid, C.H. and Brown, E.N. (1999), "A Probability Model for Saltatory Growth," In *Saltation and Stasis in Human Growth and Development: Evidence, Methods and Theory*, ed. M. Lampl, London: Smith-Gordon, 121-131.
- Seilkop, S. (1995), "The Effect of Body Weight on Tumor Incidence and Carcinogenicity Testing in B6C3F1 Mice and F344 Rats," *Fundamental and Applied Toxicology*, 24, 247-259.

- Smith, M., and Kohn, R. (1996), "Nonparametric Regression using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317-344.
- Staniswalis, J.G., and Lee, J.J. (1998), "Nonparametric Regression Analysis of Longitudinal Data," *Journal of the American Statistical Association*, 93, 1403-1418.
- Wu, H.L. and Zhang, J.T. (2002), "Local Polynomial Mixed-Effects Models for Longitudinal Data," *Journal of the American Statistical Association*, 97, 883-897.
- Zellner, A. (1962), "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregate Bias," *Journal of the American Statistical Association*, 57, 348-368.

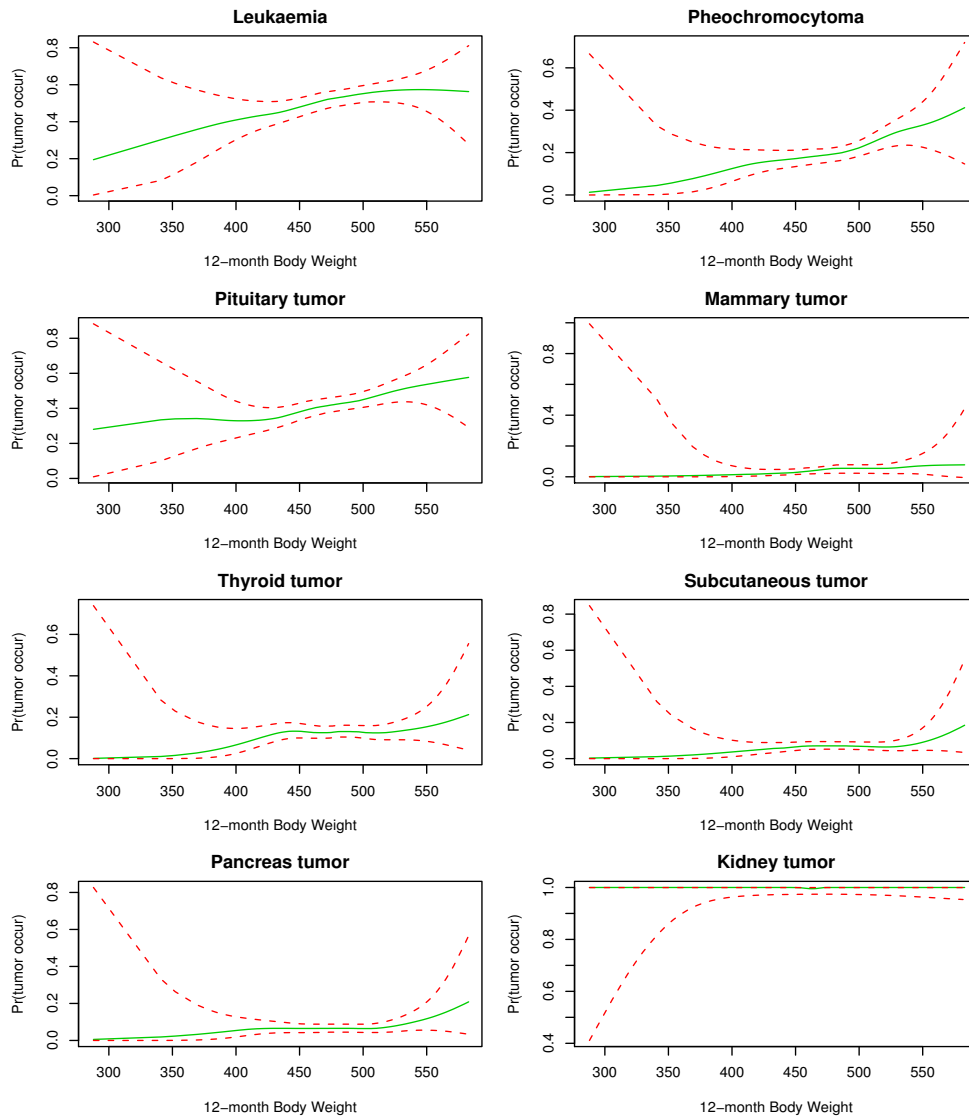


Figure 1: Estimated probability of developing a tumor of each type according to 52 week body weight. The solid line is the unconstrained frequentist estimate from a smoothing spline GAM model, fitted in S-PLUS using `gam()`, and the dashed lines are 95% confidence limits.

Table 1: Posterior probabilities of getting the tumor with the mean body weight.  $(e, f)$  are the hyperparameters for the prior of  $\delta$ .

Tumor type	$(e, f)$	$\Pr(H_{1j} Data)$	Range
<i>Leukaemia</i>	(1.50, 30)	0.531	(0.481, 0.581)
	(0.75, 15)	0.540	(0.473, 0.588)
	(3.00, 60)	0.527	(0.484, 0.577)
	(0.75, 30)	0.530	(0.479, 0.580)
	(3.00, 30)	0.538	(0.475, 0.589)
<i>Pheochromocytoma</i>	(1.50, 30)	0.188	(0.158, 0.224)
	(0.75, 15)	0.185	(0.153, 0.230)
	(3.00, 60)	0.187	(0.162, 0.221)
	(0.75, 30)	0.190	(0.159, 0.222)
	(3.00, 30)	0.192	(0.155, 0.229)
<i>Pituitary</i>	(1.50, 30)	0.412	(0.363, 0.463)
	(0.75, 15)	0.417	(0.354, 0.466)
	(3.00, 60)	0.410	(0.368, 0.459)
	(0.75, 30)	0.415	(0.367, 0.464)
	(3.00, 30)	0.413	(0.356, 0.469)
<i>Mammary</i>	(1.50, 30)	0.046	(0.039, 0.054)
	(0.75, 15)	0.044	(0.034, 0.057)
	(3.00, 60)	0.047	(0.041, 0.051)
	(0.75, 30)	0.044	(0.042, 0.052)
	(3.00, 30)	0.048	(0.036, 0.050)
<i>Thyroid</i>	(1.50, 30)	0.126	(0.106, 0.147)
	(0.75, 15)	0.123	(0.101, 0.149)
	(3.00, 60)	0.125	(0.111, 0.144)
	(0.75, 30)	0.124	(0.112, 0.146)
	(3.00, 30)	0.127	(0.100, 0.151)
<i>Subcutaneous</i>	(1.50, 30)	0.069	(0.057, 0.081)
	(0.75, 15)	0.071	(0.053, 0.086)
	(3.00, 60)	0.068	(0.061, 0.077)
	(0.75, 30)	0.068	(0.059, 0.079)
	(3.00, 30)	0.072	(0.055, 0.085)
<i>Pancreas</i>	(1.50, 30)	0.066	(0.052, 0.078)
	(0.75, 15)	0.066	(0.049, 0.080)
	(3.00, 60)	0.067	(0.055, 0.072)
	(0.75, 30)	0.068	(0.051, 0.075)
	(3.00, 30)	0.065	(0.048, 0.081)
<i>Kidney</i>	(1.50, 30)	1.00 <sup>a</sup>	(0.994, 1.00 <sup>a</sup> )
	(0.75, 15)	0.997	(0.993, 1.00 <sup>a</sup> )
	(3.00, 60)	1.00 <sup>a</sup>	(0.997, 1.00 <sup>a</sup> )
	(0.75, 30)	0.998	(0.996, 0.999)
	(3.00, 30)	0.995	(0.992, 1.00 <sup>a</sup> )

<sup>a</sup> Values > 0.999

Table 2: Posterior estimates of the threshold body weight(grams), which is the lowest weight such that there is an increased incidence relative to animals at the bottom of the weight range.

The percentile is the percentile of the estimate in the empirical distribution.

Threshold body weight	Mean	SD	95% C. I.	Percentile
$b_{leukaemia}$	321.75	17.94	(273.53, 366.81)	0.08%
$b_{pheochromocytoma}$	374.15	17.30	(332.76, 391.24)	0.41%
$b_{pituitary}$	326.67	16.14	(287.90, 378.31)	0.08%
$b_{mammary}$	383.29	19.88	(340.32, 411.25)	0.50%
$b_{thyroid}$	363.23	16.60	(339.65, 403.81)	0.25%
$b_{subcutaneous}$	392.08	19.53	(359.80, 437.94)	0.91%
$b_{pancreas}$	398.83	18.58	(366.41, 442.25)	1.49%
$b_{kidney}$	304.18	16.53	(263.78, 317.57)	0.08%

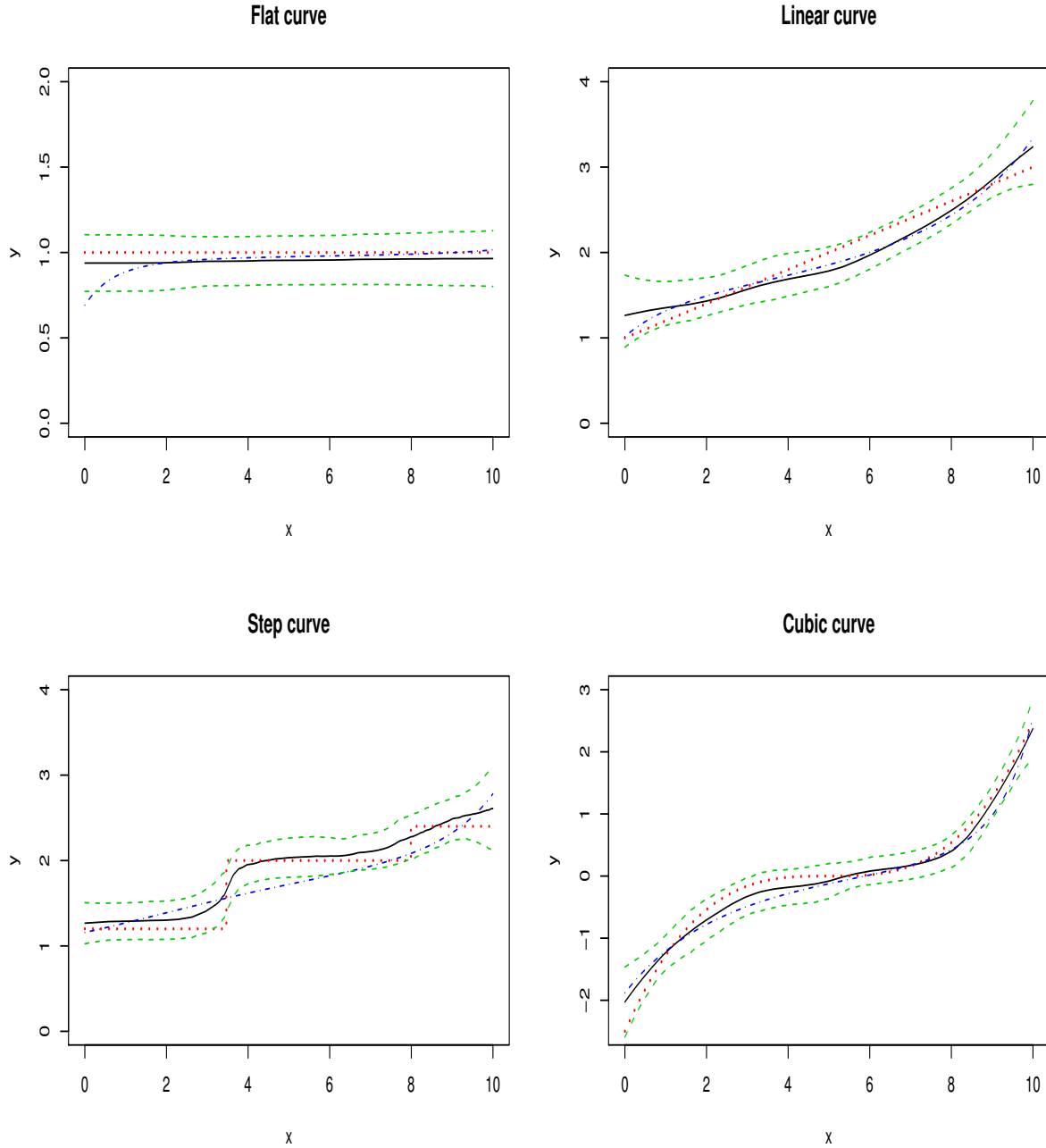


Figure 2: Estimates for univariate outcome data simulated under 1) flat curve:  $f(x) = 1$ ; 2) linear curve:  $f(x) = 1 + 0.2x$ ; 3) step curve:  $f(x) = 1.2 + 0.8I_{x>3.5} + 0.4I_{x>8}$ ; 4) cubic curve:  $f(x) = 0.02(x - 5)^3$ . Dotted line denotes true curve, solid line denotes posterior mean based on our method, dashed lines denote 95% pointwise credible intervals, and dashed-dotted line denotes the Ramsay (1998) estimates.

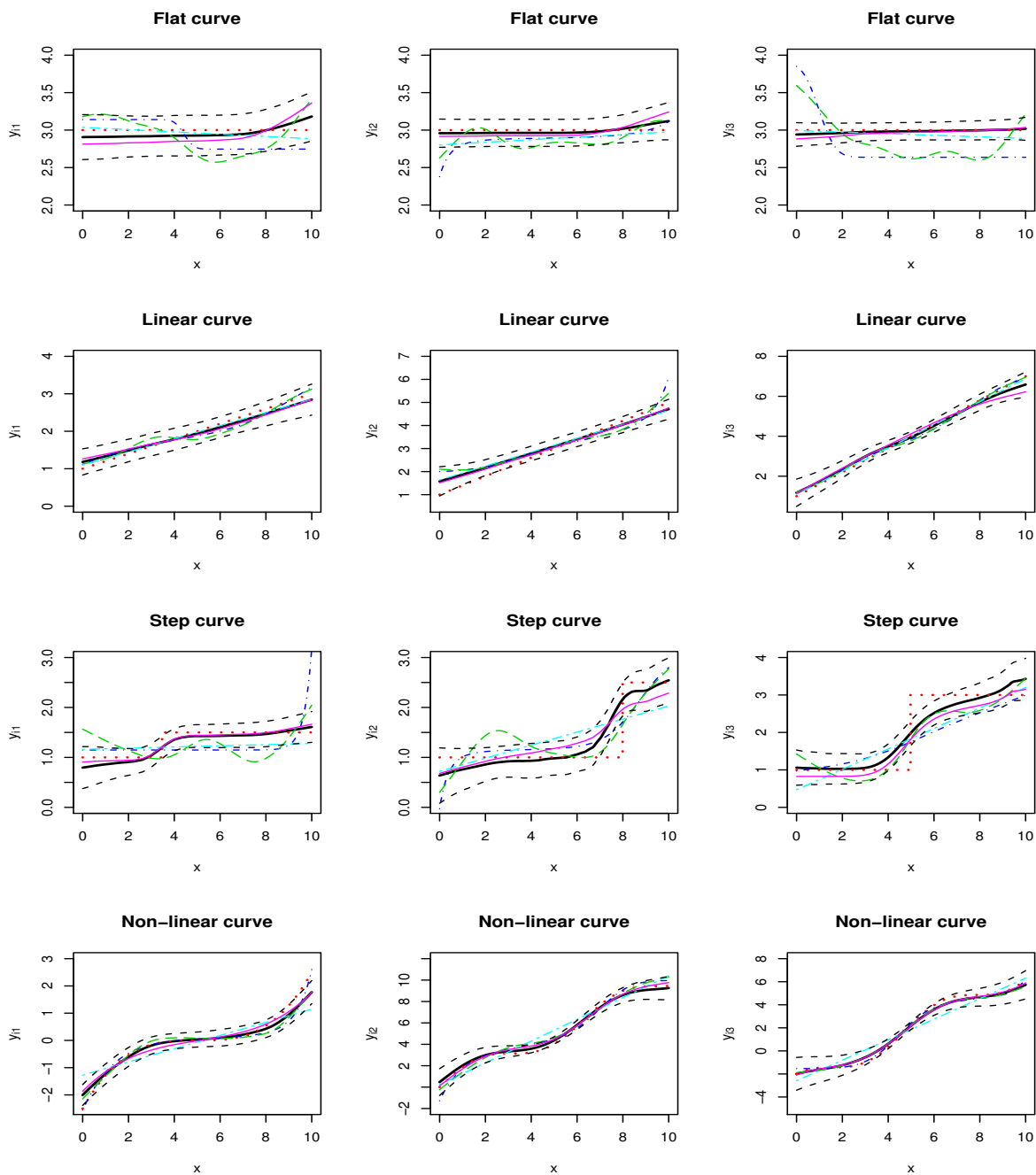
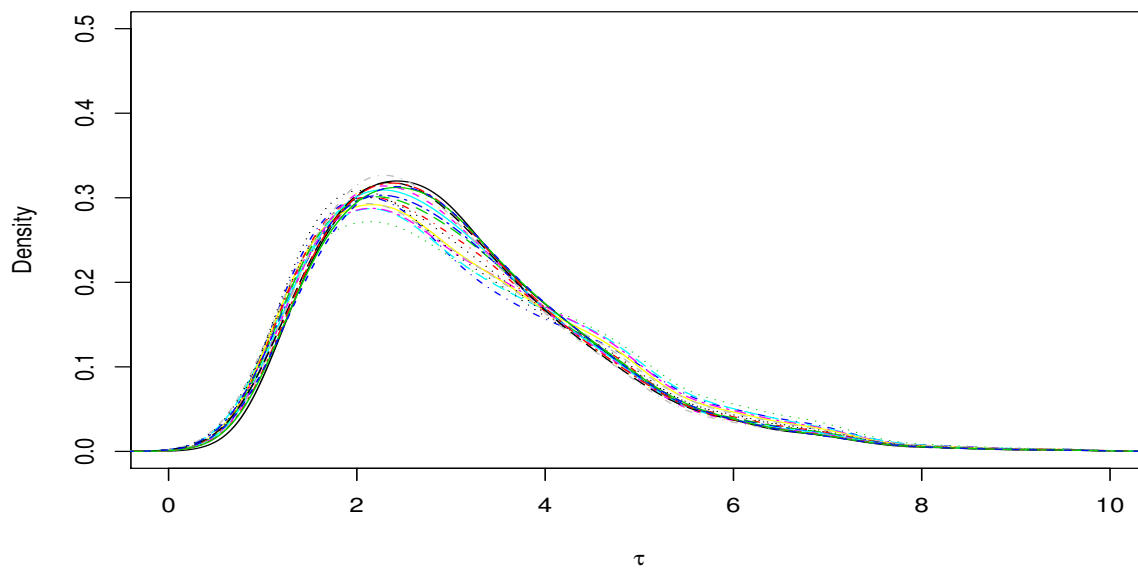


Figure 3: Estimates for multivariate outcome data simulated under four different curves. Dotted line denotes true curve, bold-solid line denotes posterior mean based on our multivariate method, thin-solid line denotes posterior mean based on our method without SUR, short-dashed lines denote 95% pointwise credible intervals, long-dashed line denotes GAM estimates, long-dashed-dotted line denotes LM estimates, and short-dashed-dotted line denotes the Ramsay (1998) estimates.



**Cubic curve**

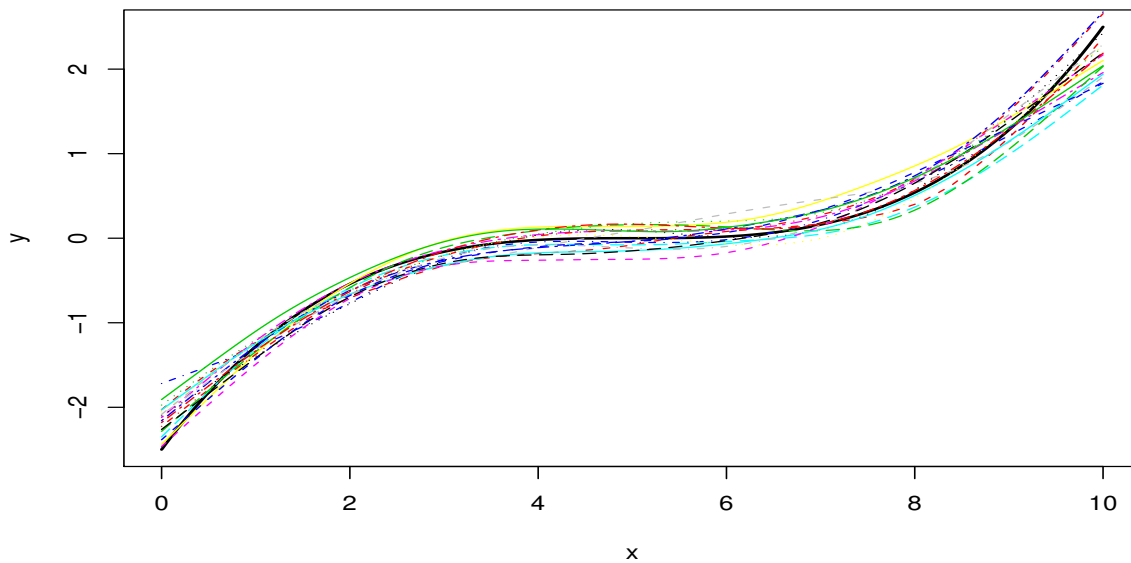


Figure 4: Sensitivity to the prior specification for  $\tau$ . The upper plot shows the posterior densities for  $\tau$  based on 20 different values of the first hyperparameter ranging from 1 to 4. The lower plot shows the corresponding curves for these hyperparameters, where the bold solid line denotes the true curve.

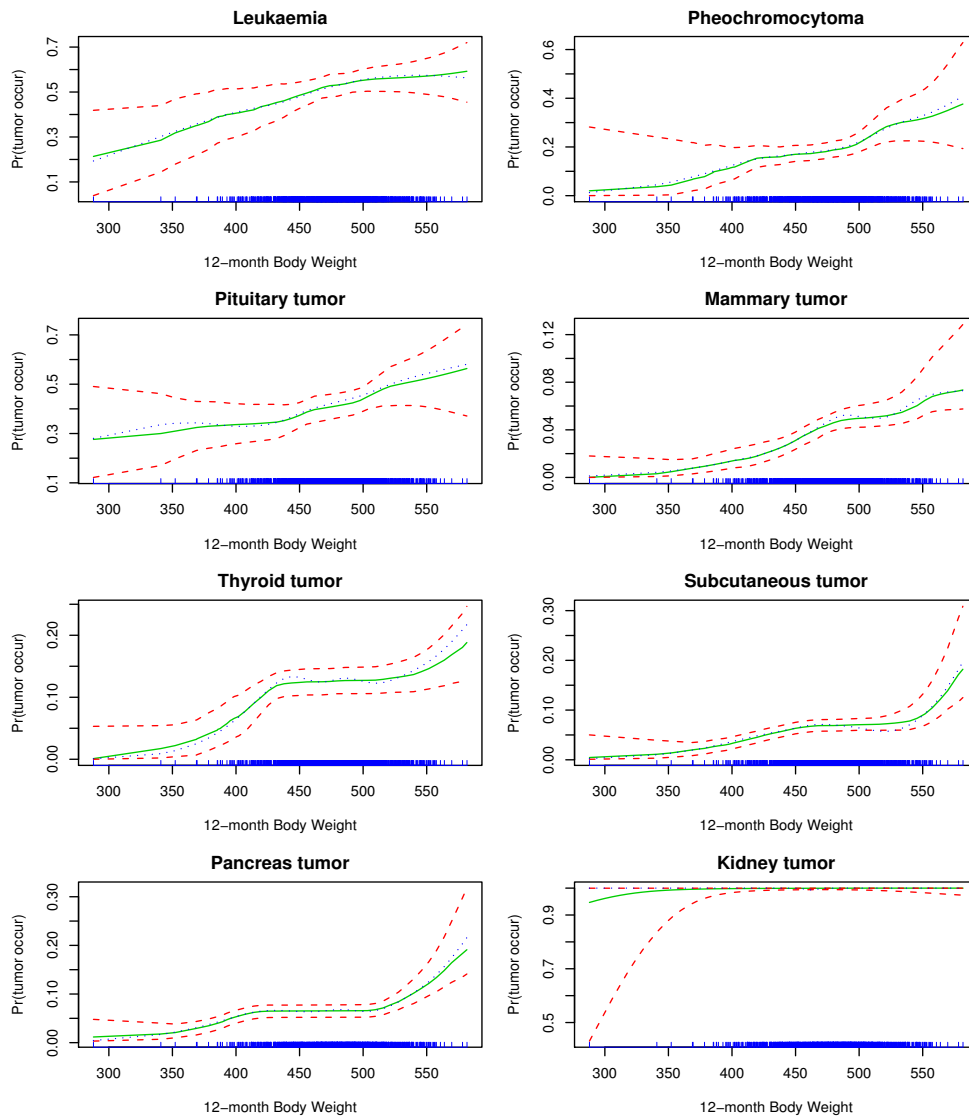


Figure 5: Estimated tumor-specific posterior means (solid line) and 95% pointwise credible intervals (dashed line) from our Bayesian isotonic regression analysis together with the estimates (dotted line) from a simple frequentist GAM analysis. Predictor values (body weights) are shown as a rug plot along the bottom of each plot.

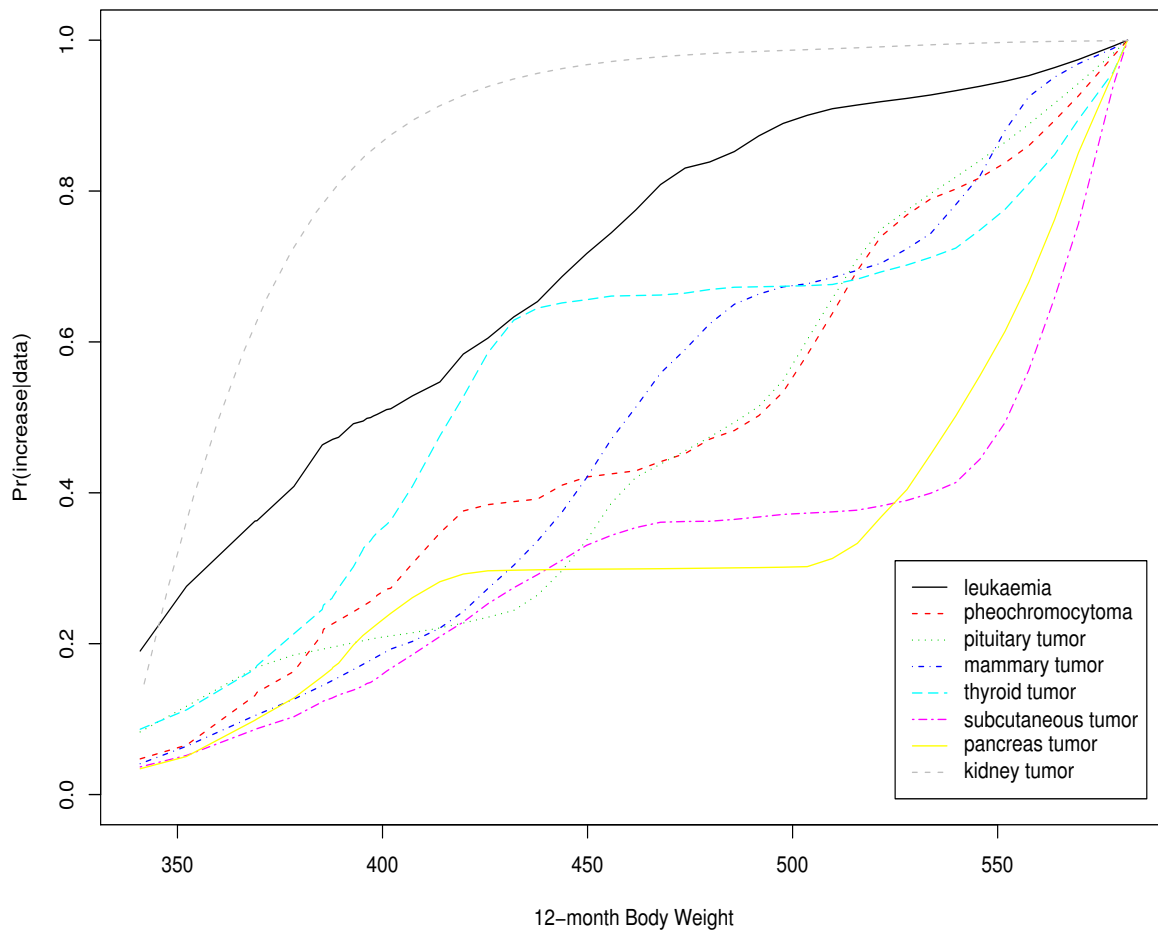


Figure 6: Estimated probability of increase of tumor incidence