Bayesian Inference of Chemical Mixtures in Risk Assessment Incorporating the Hierarchical Principle

Debamita KUNDU, Sungduk KIM, and Paul S. ALBERT*

Abstract

Analyzing health effects associated with exposure to environmental chemical mixtures is a challenging problem in epidemiology, toxicology, and exposure science. In particular, when there are a large number of chemicals under consideration it is difficult to estimate the interactive effects without incorporating reasonable prior information. Based on substantive considerations, researchers believe that true interactions between chemicals need to incorporate their corresponding main effects. In this paper, we use this prior knowledge through a shrinkage prior that a *priori* assumes an interaction term can only occur when the corresponding main effects exist. Our initial development is for logistic regression with linear chemical effects. We extend this formulation to include non-linear exposure effects and to account for exposure subject to detection limit. We develop an MCMC algorithm using a shrinkage prior that shrinks the interaction terms closer to zero as the main effects get closer to zero. We examine the performance of our methodology through simulation studies and illustrate an analysis of chemical interactions in a case-control study in cancer.

KEYWORDS AND PHRASES: Chemical mixture, Interaction, Shrinkage, Collapsed Gibbs.

1. INTRODUCTION

Assessing the health effects of environmental chemical mixtures is an important challenge in environmental epidemiology. It is quite difficult to find the relationship between exposure variables and a health outcome when the effects may be non-linear, interaction effects are present and may be subject to detection limits. In the past decade, there have been numerous approaches for analyzing this type of data. Hwang et al. [12] and Zhang et al. [21] proposed a latent class approach that links the exposure profile with disease severity with latent variables. Bobb et al. [3] employed Bayesian kernel machine regression that allows for flexible non-linear estimation of chemical effects on disease outcomes. Further, Carrico et al. [4] proposed a weighted quantile sum approach that generalizes using cumulative chemical exposure for predicting disease outcomes. Herring et al. [11] proposed a Bayesian regression approach that estimates linear main and interactive effects using a Dirichlet process prior and incorporates detection limits for the chemical exposures.

In this paper, we propose a Bayesian approach that incorporates non-linear exposure, interaction effects, and detection limits in a flexible way that does not require parametric assumptions on the exposure distributions. When the number of parameters are moderately large, the inclusion of all pairwise interaction terms results in sparsity that in turn results in poor estimation with maximum-likelihood estimation. We propose a shrinkage prior approach for estimating

interactions. First, we investigate a shrinkage prior where interactions are treated the same as main effects, and no relationship between the two is incorporated. In a second approach, we propose a shrinkage prior that incorporates a relationship between the interaction and main effects to increase performance in sparse data situations. The hierarchical principle in linear models specifies that interactions will only be examined in situations where the corresponding main effects are sizable [15]. Models with interactions without main effects place restrictions on the parameter space that are not natural. We will show how incorporating this additional structure will provide efficiency gain when studying the interactions in chemical mixtures. Additionally, we show how to extend the shrinkage prior approach to situations where multiple parameters are used to model the effect of each chemical component on disease risk. This later extension can be used to model non-linear exposure effects as well as to provide a flexible approach for dealing with detection limits in mixtures. We propose a shrinkage prior approach that incorporates the hierarchical principle for the estimation of interactions for linear and nonlinear exposures as well as for exposures subject to detection limits.

In our paper, we describe our methodology in detail in Section 2. Next in Section 3, we discuss our prior specifications and posterior computations. In Section 4 we describe how to extend our model from linear where the relationship of an exposure has a single parameter to the situation where multiple parameters are associated with that exposure (e.g., nonlinear exposures or detection limits). In Section 5 we show the efficiency of our proposed method by simulation

*Corresponding author.

studies and results. Finally, we applied our methodology to the NCI-SEER NHL study and described our findings in Section 6.

2. MODEL

Let $Y = (Y_1, Y_2, \ldots, Y_N)'$ denotes the binary health response for N individuals and $\mathbf{X_i} = (X_{i1}, X_{i2}, \ldots, X_{ip})'$ be the corresponding p-dimensional vector of continuous chemical exposures. For p chemicals we have $p(p-1)/2$ two-way interactions. We consider a logistic regression model with linear effects in their corresponding interactions of the following form:

$$\text{logit}\, P(Y_i = 1|\mathbf{X_i}) = \mathbf{U_i}'\boldsymbol{\alpha} + \sum_{j=1}^{p} X_{ij}\beta_j^*$$
$$+ \sum_{j=1}^{p} \sum_{k=j+1}^{p-1} X_{ij}X_{ik}\gamma_{jk}^*, \quad i = 1, 2, \ldots, N, \tag{2.1}$$

where logit $a = \log \frac{a}{1-a}$, $\mathbf{U_i}$ denotes q-dimensional covariate vector which includes an intercept, $\boldsymbol{\alpha}$ is the corresponding regression coefficient vector, β_j^* denotes the main effect regression coefficient of the jth chemical, and γ_{jk}^* denotes the interaction effect regression coefficient of the jth and kth chemicals. We consider a latent variable approach [1] and approximate equation (2.1) using a robit link [14]. Let's consider $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)'$ be N-dimensional latent vector such that

$$Y_i = \begin{cases} 1 & \omega_i > 0, \\ 0 & \text{otherwise}, \end{cases} \tag{2.2}$$

where, $\omega_i = \mathbf{U_i}'\boldsymbol{\alpha} + \sum_{j=1}^{p} X_{ij}\beta_j^* + \sum_{j=1}^{p}\sum_{k=j+1}^{p-1} X_{ij}X_{ik}\gamma_{jk}^* + \epsilon_i$. If $\epsilon_i \sim F_{t_v}$, where F_{t_v} is a cumulative distribution function of student t-distribution with v degrees of freedom, it is called *robit(v)* regression [13], i.e.

$$P(Y_i = 1|\boldsymbol{\alpha}, \boldsymbol{\beta^*}, \boldsymbol{\gamma^*}) = 1 - P(Y_i = 0|\boldsymbol{\alpha}, \boldsymbol{\beta^*}, \boldsymbol{\gamma^*})$$
$$= F_{t_v}\left(\mathbf{U_i}'\boldsymbol{\alpha} + \sum_{j=1}^{p} X_{ij}\beta_j^* + \sum_{j=1}^{p}\sum_{k=j+1}^{p-1} X_{ij}X_{ik}\gamma_{jk}^*\right), \tag{2.3}$$

where $\beta^* = (\beta_1^*, \beta_2^*, \ldots, \beta_p^*)$ and $\gamma^* = (\gamma_{11}^*, \gamma_{12}^*, \ldots, \gamma_{p(p-1)/2}^*)$. As $v \to \infty$, the *robit(v)* model becomes the probit regression model. Liu et al. [14] suggested that the *robit* link with $v = 7$ degrees of freedom closely approximates the *logit* link with $\beta_j = \beta_j^*/1.5484$ and $\gamma_{jk} = \gamma_{jk}^*/1.5484$. Moreover, we use the fact that the t-distribution can be represented as a scale mixture of normal distribution by introducing a mixing variable λ_i, such that $\epsilon_i|\lambda_i \sim N(0, \frac{1}{\lambda_i})$ and $\lambda_i \sim G(\frac{v}{2}, \frac{v}{2})$, where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 and $G(c_1, c_2)$ denotes the gamma distribution with mean c_1/c_2 and variance c_1/c_2^2 to formulate the likelihood. For simplicity, let

consider for the ith individual the interaction term between two exposure variables X_{ij} and X_{ik} is defined by $Z_{i_{jk}} = X_{ij}X_{ik}$ and $\boldsymbol{Z_i} = (Z_{i_{11}}, Z_{i_{12}}, \ldots, Z_{i_{p(p-1)/2}})'$. Hence $\omega_i|\lambda_i \sim N(\mathbf{U_i}'\boldsymbol{\alpha} + \boldsymbol{X_i'\beta} + \boldsymbol{Z_i'\gamma}, \frac{1}{\lambda_i})$ and $\lambda_i \sim G(\frac{v}{2}, \frac{v}{2})$ where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ and $\gamma = (\gamma_{11}, \gamma_{12}, \ldots, \gamma_{p(p-1)/2})$. From equations (2.1) and (2.2) the complete data likelihood can be written as:

$$\mathcal{L}(Y|X, w)$$
$$= \prod_{i=1}^{N} \left[y_i 1\{w_i > 0\} + (1 - y_i)1\{w_i \leq 0\}\right]$$
$$\times (2\pi)^{-\frac{1}{2}} \lambda_i^{\frac{1}{2}} \exp\left(-\frac{\lambda_i}{2}\left(w_i - \boldsymbol{U_i'\alpha} - \boldsymbol{X_i'\beta} - \boldsymbol{Z_i'\gamma}\right)^2\right)$$
$$\times \frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \lambda_i^{\frac{v}{2}-1} \exp\left(-\frac{\lambda_i v}{2}\right). \tag{2.4}$$

3. PRIOR & POSTERIOR DISTRIBUTION

For linear models with interactions, the hierarchical principle implies that interactions should only be included if the corresponding main effects are non-zero [7, 9]. Hence, we consider a dependence structure between the main and interaction effects such that the inclusion of interaction effects depends on the inclusion of the corresponding main effects. To this end, we consider the following prior distributions:

$$\beta_j \sim N\left(0, \frac{1}{a\eta_j}\right), \ \gamma_{jk} \sim N\left(0, \frac{1}{b\eta_j\eta_k\theta_{jk}}\right),$$
$$\eta_j \sim G(a_1, b_1), \ \theta_{jk} \sim G(a_2, b_2). \tag{3.1}$$

In the prior distribution in equation (3.1), a controls the global shrinkage towards the origin for the main effect regression coefficient β_j, $j = 1, 2, \ldots, p$, and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_p)'$ is the predictor specific local shrinkage parameter that allow deviations in the degree of shrinkage between predictors. a_1, a_2 defines the shape parameter and b_1, b_2 defines the scale parameter of the gamma distributions. We consider a $G(a_1 = 1, b_1 = 1)$ distribution as a prior choice for η_j with mean and variance 1 to induce some variability between the η_j's. In this formulation, larger values of η_j's will induce more shrinkage towards zero, while smaller values of η_j result in minimum shrinkage to zero. This specification is based on global-local shrinkage framework [17], where typical recommendation is to consider a heavy tail distribution for the local shrinkage parameter to avoid overshrinking large signals, and the global shrinkage parameter should have substantial mass near zero. Similarly, for the interaction effect regression coefficients γ_{jk}, the global shrinkage parameter b shrinks all parameter towards zero. In contrast, the predictor specific local shrinkage parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{p(p-1)/2})'$ captures the interaction specific shrinkage effects. We consider a $G(a_2 = 1, b_2 = 1)$ prior for θ_{jk}. Note that, to share the information between main and interaction effects, the prior variance of γ_{jk} is also dominated by the term $\eta_j\eta_k$. The parameter γ_{jk} is shrunk to zero

if at least one of η_j, η_k or θ_{jk} is large. Consistent with the hierarchical principle, an interaction term will tend to be small if either β_j or β_k is small or θ_{jk} is large. Furthermore, we also consider a $G(1,1)$ prior on both global shrinkage parameter a and b, and a vague prior $N(0, 10^2)$ for the regression coefficient α_j. The posterior distribution based on the complete data is given by

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{w}|\mathbf{X}, \mathbf{Y})$$

$$\propto \prod_{i=1}^{N}\left[y_i 1\{w_i > 0\} + (1 - y_i)1\{w_i <= 0\}\right]$$

$$\times \lambda_i^{\frac{1}{2}}\exp\left(-\frac{\lambda_i}{2}\left(w_i - \boldsymbol{U_i'\alpha} - \boldsymbol{X_i'\beta} - \boldsymbol{Z_i'\gamma}\right)^2\right) \times \lambda_i^{\frac{v}{2}-1}$$

$$\times \exp\left(-\frac{\lambda_i v}{2}\right) \times |\Lambda|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\beta'}\Lambda^{-1}\boldsymbol{\beta}\right) \times |\Omega|^{-\frac{1}{2}}$$

$$\times \exp\left(-\frac{1}{2}\boldsymbol{\gamma'}\Omega^{-1}\boldsymbol{\gamma}\right) \times |\Psi|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\alpha'}\Psi^{-1}\boldsymbol{\alpha}\right)$$

$$\times \left(\prod_{j=1}^{p}\eta_j^{a_1-1}\exp[-\eta_j b_1]\right)$$

$$\times \left(\prod_{j=1}^{p-1}\prod_{k=j+1}^{p}\theta_{jk}^{a_2-1}\exp[-\theta_{jk}b_2]\right)$$

$$\times \left(a^{a_3-1}\exp[-ab_3]\right)\left(b^{a_4-1}\exp[-bb_4]\right). \quad (3.2)$$

In equation (3.2), we define $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$, $\boldsymbol{\gamma} = (\gamma_{12}, \gamma_{13}, \ldots, \gamma_{p(p-1)})'$, $\Lambda = \text{diag}(1/a\eta_1, 1/a\eta_2, \ldots, 1/a\eta_p)$, $\Omega = \text{diag}(1/b\eta_1\eta_2\theta_{12}, 1/b\eta_1\eta_3\theta_{13}, \ldots, 1/b\eta_{p-1}\eta_p\theta_{(p-1)p})$ and $\Psi = 10^2 I_q$, where I_q represents the $q \times q$ order identity matrix.

The proposed methodology can be easily extended to incorporate other prior distributions such as horseshoe, Cauchy, and Dirichlet-Laplace prior.

3.1 Computational Development

We present a detailed development of the Markov chain Monte Carlo (MCMC) sampling algorithm. The conditional posterior distributions are derived from the equation (3.2) and inference is performed through MCMC methods. We define $X^{N \times p} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_N})'$, $Z^{N \times \frac{p(p-1)}{2}} = (\boldsymbol{Z_1}, \boldsymbol{Z_2}, \ldots, \boldsymbol{Z_N})'$. Based on the joint posterior distribution, we derive the full conditional posterior distribution of all parameters and thus get the Gibbs samples by iterating the following sampling steps:

1. We sample w_i from the conditional posterior distribution $\omega_i|\lambda_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\theta}, a, b$, which is a truncated normal distribution with mean $\boldsymbol{U_i'\alpha} + \boldsymbol{X_i'\beta} + \boldsymbol{Z_i'\gamma}$, variance λ_i^{-1}, for all $i = 1, 2, \ldots, N$.
2. For $i = 1, 2, \ldots, N$, sample λ_i independently from its conditional posterior distribution $\lambda_i|\omega_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\theta}, a, b$ which follows $G(\frac{1+v}{2}, \frac{1}{2}(v + (w_i - \boldsymbol{U_i'\alpha} - \boldsymbol{X_i'\beta} - \boldsymbol{Z_i'\gamma})^2))$.
3. Sample the covariates $\boldsymbol{\alpha}$ from its conditional multivariate normal posterior distribution $\text{MVN}(\mu_\alpha, \Sigma_\alpha)$,

where the posterior mean is $\boldsymbol{\mu_\alpha} = (U'\Sigma^{-1}U + \Psi^{-1})^{-1}U'\Sigma^{-1}(\boldsymbol{w} - \boldsymbol{X'\beta} - \boldsymbol{Z'\gamma})$, posterior variance $\Sigma_\alpha = (U'\Sigma^{-1}U + \Psi^{-1})^{-1}$ and $\Sigma = \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_N^{-1})$.
4. We sample $\boldsymbol{\beta}$ from its conditional multivariate normal posterior distribution $\text{MVN}(\mu_\beta, \Sigma_\beta)$, where posterior mean $\boldsymbol{\mu_\beta} = (X'\Sigma^{-1}X + \Lambda^{-1})^{-1}X'\Sigma^{-1}(\boldsymbol{w} - \boldsymbol{U'\alpha} - \boldsymbol{Z'\gamma})$ and posterior variance $\Sigma_\beta = (X'\Sigma^{-1}X + \Lambda^{-1})^{-1}$.
5. Sample $\boldsymbol{\gamma}$ from its conditional posterior distribution $\text{MVN}(\mu_\gamma, \Sigma_\gamma)$, where $\boldsymbol{\mu_\gamma} = (Z'\Sigma^{-1}Z + \Omega^{-1})^{-1}Z'\Sigma^{-1}(\boldsymbol{w} - \boldsymbol{U'\alpha} - \boldsymbol{X'\beta})$ and $\Sigma_\gamma = (Z'\Sigma^{-1}Z + \Omega^{-1})^{-1}$.
6. For $j = 1, 2, \ldots, p$, sample η_j independently from its conditional posterior distribution $\text{Gamma}(a_1 + \frac{p}{2}, b_1 + \frac{1}{2}(a\beta_j^2 + \sum_{k \neq j} b\eta_k\theta_{jk}\gamma_{jk}^2))$.
7. For $j = 1, 2, \ldots, p$ and $k = j + 1, 2, \ldots, p - 1$ sample θ_{jk} from the conditional posterior distribution $\text{Gamma}(a_3 + \frac{1}{2}, b_3 + \frac{1}{2}\eta_j\eta_k\gamma_{jk}^2)$.
8. Sample $a|\boldsymbol{\beta}, \boldsymbol{\eta}$ from $\text{Gamma}(a_3 + \frac{p}{2}, b_3 + \frac{1}{2}\sum_{j=1}^{p}\eta_j\beta_j^2)$.
9. Sample $b|\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\theta}$ from $\text{Gamma}(a_4 + \frac{p(p-1)}{2}, b_4 + \frac{1}{2}\sum_{j=1}^{p-1}\sum_{k=j+1}^{p}\eta_j\eta_k\theta_{jk}\gamma_{jk}^2)$.

4. MULTIPLE PARAMETERS PER EXPOSURE

Previously in Section 2, we described our method in the context of a linear exposure and outcome relationship. In many realistic settings, more than one parameter will be needed to model individual exposures in the mixtures (e.g., quadratic exposures). Hence, in this section, we extend our methodology to capture those non-linear exposure-outcome relationships using the following logistic regression model:

$$\text{logit } P(Y_i = 1|X_{ij})$$

$$= \boldsymbol{U_i'\alpha} + \sum_{j=1}^{p} g_j(X_{ij}) + \sum_{j=1}^{p}\sum_{k=j+1}^{p-1} f_{jk}(X_{ij}, X_{ik}). \quad (4.1)$$

For modeling mixtures with nonlinear exposure relationships, we can use a polynomial representation for the effect of each exposure. Polynomial effects can be incorporated in the main and interaction terms by using equation (4.1) with functions $g_j(X_{ij}) = \boldsymbol{X_{ij}'\beta_j}$ and $f_{jk}(X_{ij}, X_{ik}) = \boldsymbol{Z_{jk}'\gamma_{jk}}$ and the logistic regression can be written in the following form:

$$\text{logit } P(Y_i = 1|\boldsymbol{X_i}, \boldsymbol{Z_i})$$

$$= \boldsymbol{U_i'\alpha} + \sum_{j=1}^{p}\boldsymbol{X_{ij}'\beta_j} + \sum_{j=1}^{p}\sum_{k=j+1}^{p-1}\boldsymbol{Z_{jk}'\gamma_{jk}}, \quad i = 1, 2, \ldots, N,$$
$$(4.2)$$

where, $\boldsymbol{X_{ij}} = (X_{ij}, X_{ij}^2)'$ and $\boldsymbol{Z_{jk}} = (X_{ij}X_{ik}, X_{ij}^2X_{ik}, X_{ij}X_{ik}^2, X_{ij}^2X_{ik}^2)'$ and the regression coefficients $\boldsymbol{\beta_j} = (\beta_{j1}, \beta_{j2})'$ and $\boldsymbol{\gamma_{jk}} = (\gamma_{jk1}, \gamma_{jk2}, \gamma_{jk3}, \gamma_{jk4})'$.

Individual components of chemical mixtures may be subject to lower detection limits where the measurements are censored below these limits. Chiou et al. [6] and Ortega et al. [16] proposed a two-component exposure model for lower detection limits where the effect has one component indicating whether the exposure is above the detection limit and the other reflecting the value of the measurement if the exposure is above the detection limit. These authors showed that this parameterization does not make the unverifiable modeling assumptions inherent in treating lower detection limits as left censored in a parametric exposure distribution. We can incorporate detection limits into the mixture analysis by using equation (4.1) with:

$$
\begin{aligned}
f_{jk}&(X_{ij}, X_{ik}) \\
&= \gamma_{jk1} I(X_{ij} \geq C_j) I(X_{ik} \geq C_k) \\
&\quad + \gamma_{jk2}(X_{ij} - C_j) I(X_{ij} \geq C_j) I(X_{ik} \geq C_k) \\
&\quad + \gamma_{jk3}(X_{ik} - C_k) I(X_{ij} \geq C_j) I(X_{ik} \geq C_k) \\
&\quad + \gamma_{jk4}(X_{ij} - C_j)(X_{ij} - C_k) I(X_{ij} \geq C_j) I(X_{ik} \geq C_k).
\end{aligned}
\tag{4.3}
$$

In equation (4.3), β_{j1} is the log odds of disease at the value of the detection limit relative to the log-odds of disease below the detection limit. Further, β_{j2} is the log-odds ratio of disease for a one unit change in exposure above the detection limit. The parameter vector $\gamma_{jk.}$ measure the interactive effects between the jth and kth chemical. Specifically, γ_{jk1} measures the interactive effect of being above the detection limit on both the jth and kth chemical, γ_{jk4} measures the interactive effect of increasing X_{ij} and/or X_{ik} when both markers are above the detection limit. Finally, γ_{jk2} and γ_{jk3} are cross product interaction effects.

The shared shrinkage prior proposed in Section 3 can be extended to the multiple parameter per exposure case using the grouped shrinkage prior [5]. We show this for the non-linear formulation (4.2), but it applies more generally to other settings such as detection limits (4.3). Specifically,

$$
\begin{aligned}
\omega_i | \lambda_i &\sim N\left(U_i' \alpha + \sum_{j=1}^{p} X_{ij}' \beta_j + \sum_{j=1}^{p} \sum_{k=j+1}^{p-1} Z_{jk}' \gamma_{jk}, \frac{1}{\lambda_i}\right), \\
\lambda_i &\sim \text{Gamma}\left(\frac{v}{2}, \frac{v}{2}\right), \\
\beta_j &\sim \text{MVN}_{k_1}\left(0, \frac{\eta_j}{a} I_2\right), \\
\eta_j &\sim \text{Gamma}\left(\frac{k_1 + 1}{2}, \frac{1}{2}\right), \\
\gamma_{jk} &\sim \text{MVN}_{k_2}\left(0, \eta_j \eta_k \frac{\theta_{jk}}{b} I_4\right), \\
\theta_{jk} &\sim \text{Gamma}\left(\frac{k_2 + 1}{2}, \frac{1}{2}\right), \\
\alpha &\sim N\left(0, 10^2\right), \ a, b \sim \text{Gamma}(1, 0.1).
\end{aligned}
\tag{4.4}
$$

Here k_1 and k_2 defines the dimension of the parameters β_j and γ_{jk}, respectively. In the non-linear case we share the

information between main effect and interaction effect similarly as in linear hierarchical model in the equation (3.1). Posterior calculation follows as in Section 3.1.

5. SIMULATION STUDY AND RESULTS

We perform a series of the simulation studies to investigate the performance of the proposed methodology. We compare the proposed model with following two models.

1. **Independent Vague Prior**: We incorporate the vague prior $\beta_j, \gamma_{jk} \sim N(0, 10^2)$, where no dependence is introduced between the main and the interaction effects. This approach is approximately a maximum likelihood approach.
2. **Independent shrinkage prior**: Under this approach we do not share the information between main effects and interaction effects. We consider an independent shrinkage prior on both interaction effect and main effects regression coefficients. To that end, $\beta_j \sim N(0, \frac{1}{a\eta_j})$, $\gamma_{jk} \sim N(0, \frac{1}{b\theta_{jk}})$. Hence in this model, the inclusion of an interaction effect is not contingent on the inclusion of main effects.

5.1 Simulation for Linear Exposure Effect

We conduct a simulation study for the linear hierarchical model in equation (3.1), where data are generated using model in equation (2.2) in Section 2. For each ith individual we generate p chemical exposures $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})'$ independently from a multivariate normal distribution with mean zero and covariance matrix I_p. Thus we have our design matrix $X^{N \times p}$ for chemical exposure. Then we generate $p(p-1)/2$ interaction effects for each individual by multiplying the corresponding main effects, i.e. $Z_i = (X_{i1}X_{i2}, X_{i1}X_{i3}, \ldots, X_{ip}X_{i(p-1)})'$, where $l = 1, 2, \ldots, p(p-1)/2$. Later we scaled each of the main and interaction exposures by dividing their respective standard deviations to perform the analysis. We generate 200 datasets for each model, and for each data set we run the MCMC chain for 50,000 iterations with a burn-in of 5000 iterations. We consider every 5th sample to reduce the autocorrelation. We use those 9000 MCMC samples for posterior inferences. We consider different simulation scenarios to access the utility of our methodology in details.

- **Simulation #1:** $N = 1000$, $p = 10$: We generate data under a model consistent with the hierarchical principle. We consider $N = 1000$ individual and 10 chemicals ($p = 10$). We further consider 5 out of 10 main effects have effect size 1 and 10 out of $\binom{10}{2} = 45$ interaction terms has effect size 0.5 and rest are set to zero. The intercept is set to 0.6 to have the overall prevalence rate of approximately 50%.
- **Simulation #2:** $N = 250$, $p = 10$: In this settings, we consider a smaller sample size $N = 250$ individual and total number of chemicals $p = 10$. The main effect

Table 1. Results from simulation #1: $N = 1000$, $p = 10$.

Parameters	γ	Independent Vague			Independent shrinkage			Shared Shrinkage			Relative*
---	---	---	---	---	---	---	---	---	---	---	Efficiency
		$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	
Interactions	$\gamma_{12} = 0.5$	0.892	0.040	0.472	0.546	0.023	0.169	0.533	0.022	0.151	1.252
	$\gamma_{13} = 0.5$	0.883	0.040	0.462	0.539	0.022	0.162	0.529	0.022	0.152	1.136
	$\gamma_{14} = 0.5$	0.904	0.041	0.470	0.553	0.023	0.154	0.533	0.022	0.160	0.925
	$\gamma_{15} = 0.5$	0.914	0.040	0.494	0.563	0.022	0.179	0.528	0.021	0.152	1.387
	$\gamma_{16} = 0$	0.003	0.024	0.195	0.001	0.013	0.113	0.009	0.011	0.104	1.179
	$\gamma_{17} = 0$	-0.006	0.025	0.203	-0.004	0.014	0.123	0.005	0.011	0.098	1.575
	$\gamma_{18} = 0$	0.005	0.024	0.184	0.003	0.013	0.106	0.009	0.011	0.091	1.357
	$\gamma_{19} = 0$	0.000	0.024	0.174	-0.003	0.013	0.102	0.000	0.011	0.097	1.104
	$\gamma_{1,10} = 0$	0.003	0.024	0.199	-0.001	0.013	0.114	-0.005	0.011	0.103	1.225
	$\gamma_{23} = 0.5$	0.869	0.039	0.448	0.532	0.022	0.159	0.537	0.022	0.155	1.053
	$\gamma_{24} = 0.5$	0.907	0.041	0.486	0.559	0.023	0.175	0.524	0.021	0.152	1.325
	$\gamma_{25} = 0.5$	0.875	0.039	0.464	0.536	0.022	0.172	0.538	0.022	0.154	1.248
	$\gamma_{26} = 0$	0.010	0.024	0.190	0.005	0.013	0.110	0.004	0.011	0.088	1.563

*Relative efficiency=Ratio of MSE of independent shrinkage prior and the MSE of shared shrinkage prior.
The geometric mean efficiency gain across all 45 interaction terms is 1.24.

Table 2. Results from simulation #2: $N = 250$, $p = 10$.

Parameters	γ	Independent Vague			Independent shrinkage			Shared Shrinkage			Relative*
---	---	---	---	---	---	---	---	---	---	---	Efficiency
		$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	
Interactions	$\gamma_{12} = 0.5$	3.949	1.546	4.300	0.563	0.068	0.284	0.565	0.068	0.273	1.082
	$\gamma_{13} = 0.5$	3.930	1.574	4.348	0.584	0.074	0.300	0.540	0.066	0.248	1.464
	$\gamma_{14} = 0.5$	3.895	1.554	4.266	0.568	0.070	0.283	0.548	0.068	0.260	1.184
	$\gamma_{15} = 0.5$	4.185	1.617	4.596	0.595	0.073	0.327	0.562	0.069	0.280	1.364
	$\gamma_{16} = 0$	-0.081	0.767	1.889	-0.013	0.040	0.208	0.005	0.032	0.169	1.515
	$\gamma_{17} = 0$	-0.073	0.716	1.672	-0.009	0.040	0.187	0.012	0.031	0.162	1.331
	$\gamma_{18} = 0$	0.087	0.760	1.784	0.007	0.041	0.210	-0.001	0.031	0.152	1.910
	$\gamma_{19} = 0$	0.038	0.704	1.782	0.003	0.040	0.216	-0.002	0.032	0.161	1.800
	$\gamma_{1,10} = 0$	-0.021	0.759	1.801	-0.001	0.040	0.195	0.017	0.033	0.174	1.256
	$\gamma_{23} = 0.5$	4.084	1.621	4.523	0.597	0.073	0.323	0.521	0.063	0.255	1.605
	$\gamma_{24} = 0.5$	3.955	1.478	4.236	0.584	0.072	0.341	0.558	0.068	0.305	1.252
	$\gamma_{25} = 0.5$	4.077	1.669	4.431	0.589	0.070	0.295	0.540	0.067	0.278	1.126
	$\gamma_{26} = 0$	-0.110	0.759	1.873	-0.019	0.040	0.214	-0.017	0.032	0.162	1.745

*Relative efficiency=Ratio of MSE of independent shrinkage prior and the MSE of shared shrinkage prior.
The geometric mean efficiency gain across all 45 interaction terms is 1.41.

parameters $\boldsymbol{\beta}$ and interaction effect parameters $\boldsymbol{\gamma}$ are identical to simulation #1.

- **Simulation #3:** $N = 1000$**, Main effects and no interaction effects**: In this settings, we consider a sample size of $N = 1000$ individual and a total number of chemicals $p = 10$. We generate the data from a model with the same main effect parameters $\boldsymbol{\beta}$ as in simulation #1 and simulation #2, but the interaction effect parameters $\boldsymbol{\gamma}$ are set to zero, i.e, we have main effects but no interaction effects.
- **Simulation #4:** $N = 1000$**, Interactions with no main effects**: We generate data from a model with no main effects but with the same interaction effects as in simulation #1 and simulation #2.

In this section, we compare the performance of the proposed methodology with the model with an independent

vague prior and independent shrinkage model. Tables 1 and 2 show the mean standard deviation and root mean square error (RMSE) for a few of the interaction terms in simulation #1 and simulation #2 respectively. First the results show that using a vague prior result in poor estimation, particularly for smaller sample size. Secondly, incorporating the shared shrinkage prior result in a large efficiency gain relative to the independent shrinkage prior. We computed the geometric mean of the relative efficiencies comparing the independent versus the shared shrinkage across all 45 interaction parameters. Shared shrinkage showed efficiency advantages in Table 1 and Table 2 with mean efficiencies of 1.25 and 1.42, respectively. As expected, the efficiency gains were larger for the smaller sample size that has increased sparsity.

In simulation #3 (Table 3), we evaluated the situation where there were main effects but no interaction effects in the model; this situation is consistent with the shared

Table 3. Results from simulation #3: No interaction effects; $N = 1000$, $p = 10$.

| Parameters | γ | Independent Vague | | | Independent shrinkage | | | Shared Shrinkage | | | Relative* |
		$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	Efficiency
Interactions	$\gamma_{12} = 0$	−0.030	0.040	0.260	0.000	0.019	0.137	−0.008	0.013	0.102	1.804
	$\gamma_{13} = 0$	−0.055	0.039	0.272	−0.008	0.019	0.137	−0.006	0.013	0.096	2.037
	$\gamma_{14} = 0$	−0.027	0.040	0.271	0.005	0.019	0.138	−0.002	0.013	0.098	1.983
	$\gamma_{15} = 0$	−0.035	0.040	0.264	0.012	0.018	0.140	0.009	0.013	0.097	2.083
	$\gamma_{16} = 0$	0.008	0.039	0.249	−0.008	0.017	0.121	−0.004	0.011	0.082	2.177
	$\gamma_{17} = 0$	−0.009	0.038	0.264	−0.018	0.017	0.129	0.002	0.011	0.079	2.664
	$\gamma_{18} = 0$	−0.003	0.039	0.257	0.011	0.017	0.123	−0.007	0.011	0.078	2.486
	$\gamma_{19} = 0$	−0.002	0.039	0.249	−0.001	0.017	0.119	0.003	0.011	0.085	1.960
	$\gamma_{1,10} = 0$	−0.020	0.039	0.256	−0.012	0.017	0.107	0.015	0.011	0.089	1.445
	$\gamma_{23} = 0$	−0.046	0.040	0.252	−0.003	0.019	0.140	−0.012	0.013	0.094	2.218
	$\gamma_{24} = 0$	−0.003	0.041	0.261	0.010	0.019	0.131	−0.003	0.013	0.094	1.942
	$\gamma_{25} = 0$	−0.044	0.040	0.287	−0.005	0.018	0.133	0.008	0.013	0.104	1.635
	$\gamma_{26} = 0$	0.003	0.039	0.264	0.003	0.017	0.125	−0.009	0.011	0.085	2.163

*Relative efficiency=Ratio of MSE of independent shrinkage prior and the MSE of shared shrinkage prior.
The geometric mean efficiency gain across all 45 interaction terms is 2.10.

Table 4. Results from simulation #4: No main effects; $N = 1000$, $p = 10$.

| Parameters | γ | Independent Vague | | | Independent shrinkage | | | Shared Shrinkage | | | Relative* |
		$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	Efficiency
Interactions	$\gamma_{12} = 0.5$	0.745	0.021	0.307	0.579	0.015	0.154	0.559	0.015	0.138	1.245
	$\gamma_{13} = 0.5$	0.739	0.020	0.298	0.567	0.015	0.131	0.551	0.014	0.140	0.875
	$\gamma_{14} = 0.5$	0.744	0.021	0.297	0.576	0.015	0.148	0.545	0.014	0.143	1.071
	$\gamma_{15} = 0.5$	0.752	0.021	0.308	0.594	0.015	0.162	0.557	0.015	0.135	1.440
	$\gamma_{16} = 0$	−0.001	0.012	0.126	−0.007	0.009	0.095	0.002	0.007	0.084	1.279
	$\gamma_{17} = 0$	0.002	0.012	0.134	−0.001	0.009	0.096	−0.005	0.007	0.083	1.338
	$\gamma_{18} = 0$	0.000	0.012	0.139	0.004	0.009	0.095	0.007	0.007	0.077	1.522
	$\gamma_{19} = 0$	−0.011	0.012	0.119	0.003	0.009	0.095	−0.008	0.007	0.069	1.896
	$\gamma_{1,10} = 0$	0.000	0.012	0.139	−0.009	0.009	0.100	−0.006	0.007	0.074	1.826
	$\gamma_{23} = 0.5$	0.758	0.021	0.306	0.583	0.015	0.143	0.543	0.014	0.129	1.229
	$\gamma_{24} = 0.5$	0.769	0.021	0.318	0.588	0.015	0.155	0.551	0.015	0.133	1.358
	$\gamma_{25} = 0.5$	0.748	0.020	0.296	0.570	0.015	0.149	0.548	0.014	0.125	1.421
	$\gamma_{26} = 0$	0.001	0.012	0.136	0.004	0.009	0.093	−0.016	0.007	0.075	1.537

*Relative efficiency=Ratio of MSE of independent shrinkage prior and the MSE of shared shrinkage prior.
The geometric mean efficiency gain across all 45 interaction terms is 1.36.

shrinkage prior specification. Table 4 shows the results for simulation #3 where the average relative efficiency for the shared versus independent shrinkage prior was 2.1. In simulation #4, we generated data with no main effects but sizeable interaction effects. The results showed that even in the situation that does not directly correspond to the shared shrinkage prior, we found efficiency gains by using this more general prior specification; we found an average efficiency gain of 1.36 across all 45 interaction parameters (geometric mean).

5.2 Simulation for Detection Limit Model

We performed a simulation study for the multi-parameter exposure model that incorporates lower detection limits as described in Section 4. For this case, we considered $p = 10$ main effects and 5 out of these main effects have sizable effect. We need to specify two parameters for each exposure variable: (i) intercept term: $\beta_{j1}I(X_{ij} \geq C_j)$ and (ii) slope term: $\beta_{j2}I(X_{ij} \geq C_j)(X_{ij} - C_j)$. For the non-null main effects we consider the effect size $\beta_{j1} = 0.2$ and $\beta_{j1} = 0.5$ for the simulation study. Similarly, for the interaction term of any two main effects we have four parts:

1. $\gamma_{jk1}I(X_{ij} \geq C_j)I(X_{ik} \geq C_k)$
2. $\gamma_{jk2}(X_{ij} - C_j)I(X_{ij} \geq C_j)I(X_{ik} \geq C_k)$
3. $\gamma_{jk3}(X_{ik} - C_k)I(X_{ij} \geq C_j)I(X_{ik} \geq C_k)$
4. $\gamma_{jk4}(X_{ij} - C_j)(X_{ik} - C_k)I(X_{ij} \geq C_j)I(X_{ik} \geq C_k)$.

For the interactions terms that have non-null main effects we consider $\gamma_{jk4} = 0.1$ for simulation study. All other γ_{jk1}, γ_{jk2}, and γ_{jk3}'s are set to zero. Similarly as before, we generate the chemical exposure X_{ij} from a standard normal distribution. For first 5 exposure variables we consider 20% values are below the detection limit and for the rest 10% below the detection limit. Lastly, we consider the overall mean

Table 5. Results from simulation for two parameter detection limits model; $N = 1000$, $p = 10$.

Parameters	γ	Independent Vague			Independent shrinkage			Shared Shrinkage			Relative* Efficiency
		$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	$\hat{\gamma}$	sd	RMSE	
Interactions	$\gamma_{124} = 0.1$	0.016	0.509	0.717	0.164	0.199	0.450	0.141	0.115	0.341	1.741
	$\gamma_{134} = 0.1$	0.004	0.598	0.778	0.166	0.215	0.468	0.133	0.136	0.369	1.608
	$\gamma_{144} = 0.1$	−0.039	0.577	0.770	0.111	0.204	0.451	0.116	0.146	0.382	1.394
	$\gamma_{154} = 0.1$	−0.078	0.677	0.840	0.094	0.200	0.447	0.102	0.133	0.363	1.516
	$\gamma_{164} = 0$	−0.366	3.252	1.836	0.055	0.362	0.603	0.039	0.181	0.427	1.994
	$\gamma_{174} = 0$	−0.426	2.577	1.657	−0.036	0.320	0.565	0.022	0.164	0.404	1.996
	$\gamma_{184} = 0$	−0.344	1.641	1.323	−0.036	0.321	0.566	0.006	0.164	0.404	1.963
	$\gamma_{194} = 0$	−0.478	1.902	1.457	0.002	0.293	0.540	−0.006	0.163	0.403	1.795
	$\gamma_{264} = 0$	−0.423	2.123	1.514	−0.042	0.369	0.608	0.013	0.202	0.448	1.842
	$\gamma_{274} = 0$	−0.141	0.267	0.534	0.029	0.082	0.287	0.036	0.090	0.302	0.903
	$\gamma_{284} = 0$	−0.115	0.309	0.567	0.040	0.089	0.342	0.054	0.115	0.300	1.300
	$\gamma_{294} = 0$	−0.074	0.272	0.526	0.046	0.074	0.315	0.075	0.094	0.275	1.312
	$\gamma_{344} = 0$	−0.128	0.283	0.546	0.052	0.078	0.318	0.045	0.099	0.283	1.263

*Relative efficiency=Ratio of MSE of independent shrinkage prior and the MSE of shared shrinkage prior.

The geometric mean efficiency gain across all 180 interaction terms is 1.37.

$\alpha = -2.1$ to have a prevalence of 50%. We have a total of 201 parameters in this multi-parameter per exposure model. As the number of parameters are higher than linear model we consider $N = 1000$ observations for simulation study.

Table 5 shows the simulation results comparing the shared shrinkage, independent shrinkage and vague prior. Similar to the linear exposure model simulation, we show for this simulation substantial efficiency gain in using shared shrinkage for this more complex model. We also computed the geometric mean efficiency gain of shared shrinkage versus independent shrinkage across all 180 interaction parameters in this multi-parameter model. The mean efficiency gain was 1.37, reflecting substantial gains for the shared shrinkage prior in this sparse data structure.

6. NCI-SEER NHL STUDY

The NCI-SEER NHL study [18] is a population-based case-control study (508 controls & 672 cases) of non-Hodgkin lymphoma (NHL), was designed to determine the associations between exposures of chemicals/pesticides found in used vacuum cleaner bags and the risk of NHL. Different epidemiological pieces of evidence confirm that exposure to chemicals increases the risk of certain cancers in humans [20]. Often chemicals enter the household from indoor use or drift in from outdoor and may persist for months and years in carpet and cushion furniture without being degraded by sunlight, rain, and extreme temperatures. Hence carpet dust sampling provides a more objective basis for exposure assessment as it contains integrated pesticide exposure over a long period which is potentially more relevant to disease risk than recent or current exposure. In this study, the samples were collected from used vacuum cleaner bags of 672 cases in Detroit, Iowa, Los Angeles, and Seattle and were analyzed for pesticides [8]. Primarily the laboratory measurements contain missing data due to concentrations being below the minimum detection level. The "fill-in"
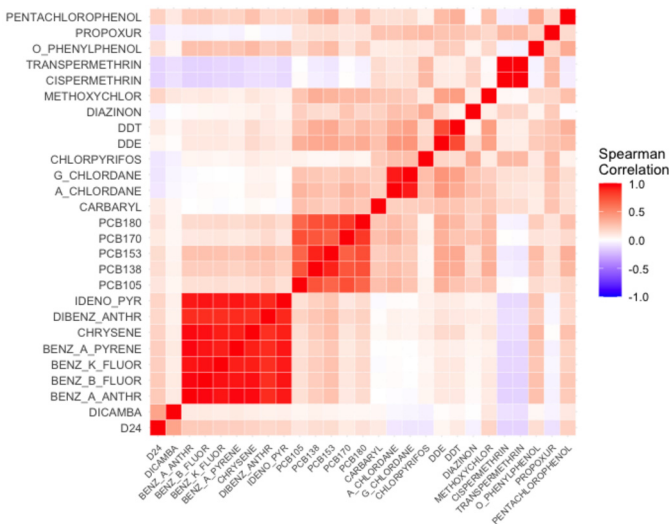

Figure 1: Correlation plot of all chemicals.

approach [10, 8] was used for imputation where, for each biomarker, measurements below the detection limit were imputed by first estimating maximum-likelihood estimators of a log-normal distribution from the left-censored data and then imputing values below the detection limit based on this distribution.

Particularly for chemicals with a high percentage of values below their detection limits, results may not be robust to misspecification of the parametric assumptions. The median percent of observations below the detection limit was 61% (across chemicals) with a range of (3% to 93%). Some of the pesticides are highly correlated (> 0.9) and skewed. Figure 1 shows the correlation between the chemicals. In that case, we chose to use only one of them for the analysis and log-transformed the exposure data. In the final data set, we have
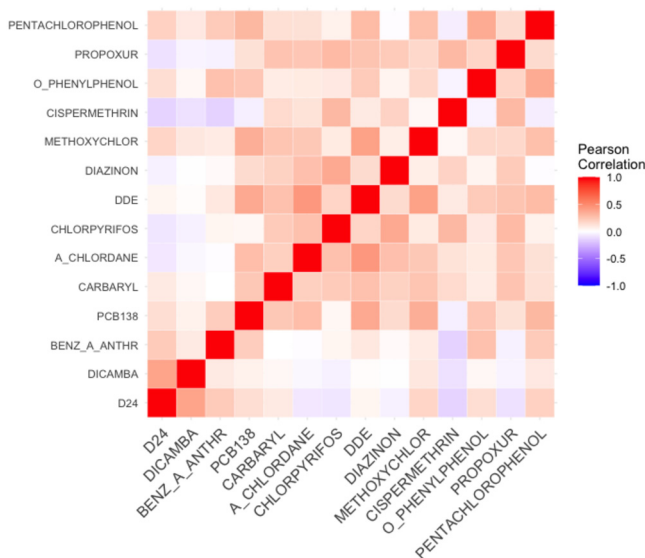
Figure 2: Correlation plot of finally selected chemicals.

a total of $p = 14$ chemical exposures on $N = 1180$ individuals (508 controls & 672 cases). Figure 2 shows the correlation between the selected chemicals for the analysis. We considered site, sex, education, and age as covariates [8] in all models for our data application. First we estimated a main effect model using a vague prior and an independent shrinkage prior. Figure 3 shows only one chemical, Diazinon, had a 95% HPD interval that excludes zero, for all other chemicals the 95% HPD interval contained zero. Figure 5 shows a random subset of interaction effects estimated under all three prior distributions. The 95% HPD intervals included zero for all estimated interactions for all priors. The intervals for the shared were narrower than the independence shrinkage, and both were narrower than the vague prior.

The proposed multi-parameter exposure model that incorporates detection limits does not make strong assumptions on the distributions and exposure effects for values below the detection limit. Figures 4 and 6 show the results from fitting the multi-parameter model. These figures show a random set of interaction effects corresponding to the slope-slope interactions (corresponding to γ_{jk4} parameters). Unlike for the linear exposure models presented, we see strong evidence for interactive effects. Figure 7 shows a contour plot obtained from the estimated terms corresponding to the interaction of D24 and Diazinon. The figure demonstrates a very strong qualitative interaction where exposure effects the risk of NHL only when an individual is exposed to both agents. (note that the log-odds ratios are near zero when either of the agents have exposure at or below their detection limit).

7. DISCUSSION

In this paper we have proposed a shrinkage prior that shares the information between the main effects and interac-

tion effects for estimating the complex relationship between chemical mixtures and disease risk. We proposed a prior that shrinks the interaction term closer to zero when there is a little evidence of a corresponding main effect. This approach is consistent with the hierarchical principle that argues that one should only look for interactions when the corresponding main effects are present. Through simulations, this approach showed sizable efficiency gain in using the shared shrinkage prior when the hierarchical principle holds. Interestingly, we saw that even when there was no main effect, the shared shrinkage prior showed efficiency over priors that did not incorporate this shared information. We presume this is due to the flexibility of the shared shrinkage prior. Bien et al. [2] propose a penalized likelihood lasso approach for incorporating the hierarchical principle for parameter estimation. However, there are advantages to using the Bayesian approach over the likelihood approach is that i) there is no need to estimate the penalization constant, and ii) estimating measures of uncertainty are more straightforward using MCMC. Bayesian kernel machine regression (BKMR, [3]) is an alternative approach that can be used to identify two-way interactions that do not impose an additive structure in the model formulation. However, this approach does not incorporate the hierarchical principle in the estimation of interaction between mixture components. Further, BKMR cannot be easily extended to flexibly incorporate detection limits as discussed in Section 4.

A major contribution of this work is introducing the shared shrinkage prior to the setting of chemical mixtures. A major analytical issue in the analysis of chemical mixtures is the presence of lower detection limits among the multiple chemicals. We proposed two parameter exposure models that incorporate these detection limits. Specifically for each chemical, we include one parameter for the effect of being above the detection limit and another parameter for the linear change (slope) when above this limit. At the expense of adding additional parameters, the proposed approach makes no assumptions about the exposure effects and chemical distributions below detection limits.

The analysis of the NHL mixture data provided interesting insights into the methodology. When using a linear exposure model that imputed values below the detection limits, we found no interactive effects using the independence or shared shrinkage prior. These analyses were based on an imputation approach where chemicals were assumed to be log-normally distributed and values were imputed in the tail of this distribution. The percent of values below the detection limits varied across chemical but was generally very high (median 61%). We would expect the inferences using imputation and assuming a linear relationship to be highly sensitive to the imposed parametric assumptions. Ortega et al. [16] showed the lack of robustness of this type of imputation for a single exposure. When we fit the two-parameter exposure model that accounts for detection limits, we found substantially more evidence for interactive effects. Most of
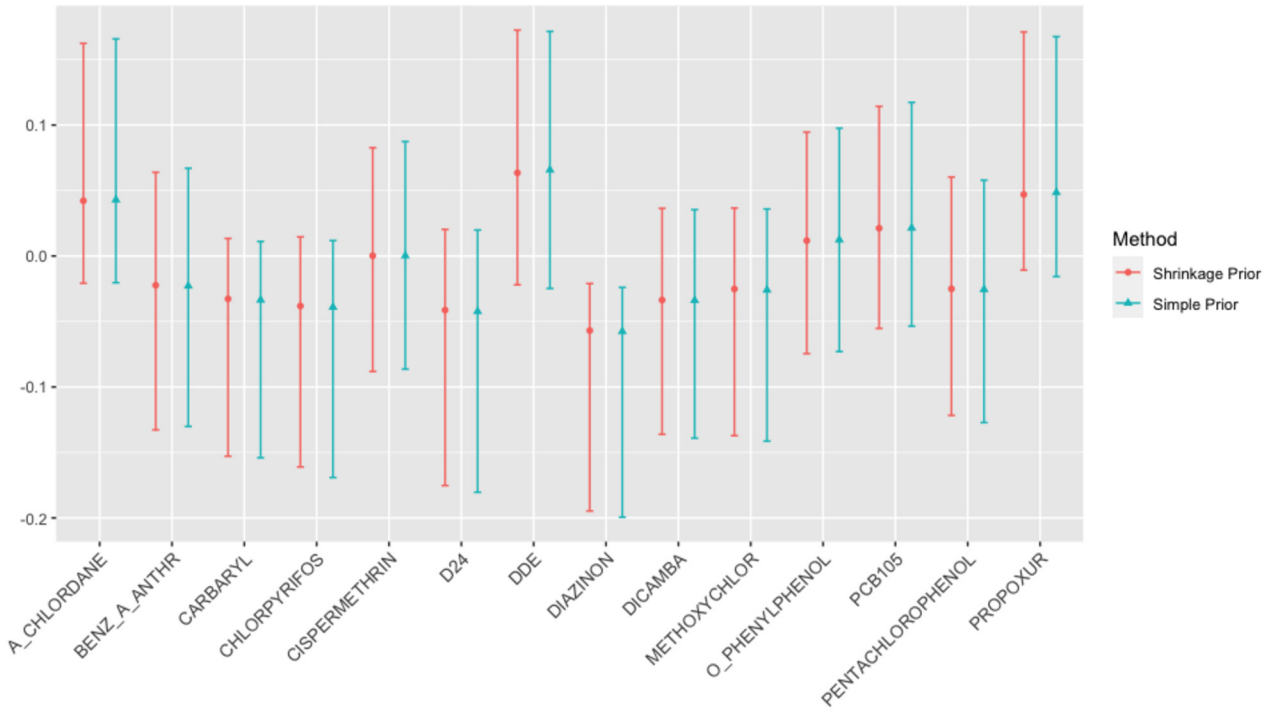
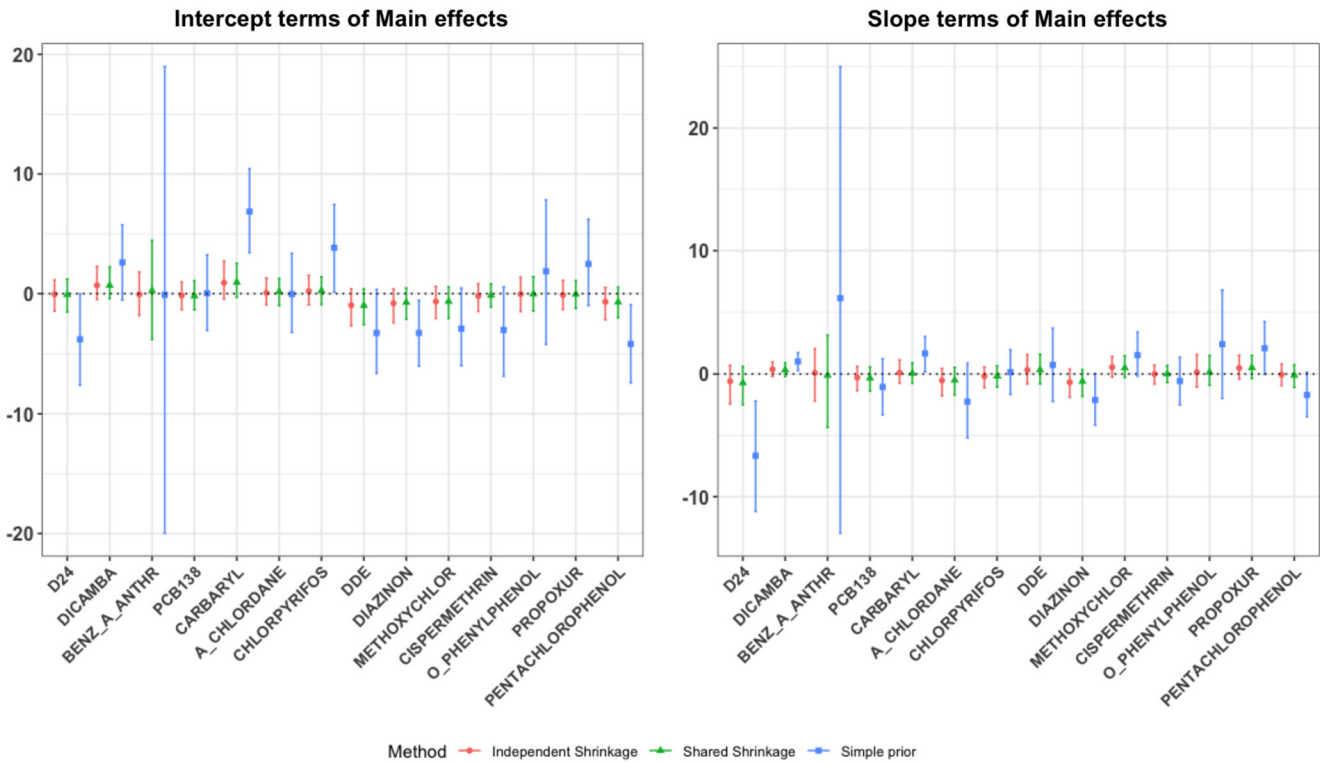Figure 3: Main effects estimation.

Figure 4: Plots for Intercept terms: $I(X_{ij} > C_{ij})$ and slopes terms: $I(X_{ij} > C_{ij})(X_{ij} - C_j)$ for the NCI-SEER NHL data set.
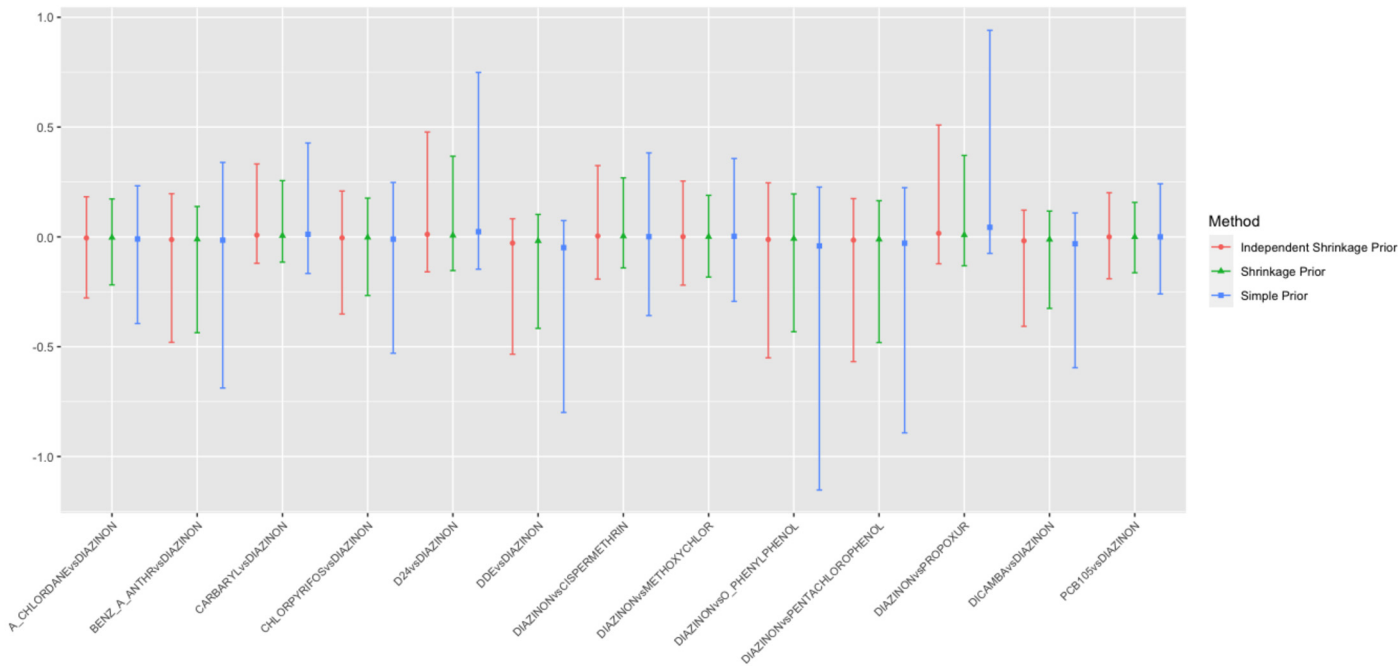
Figure 5: Interaction effect with Diazinon from linear exposure model.
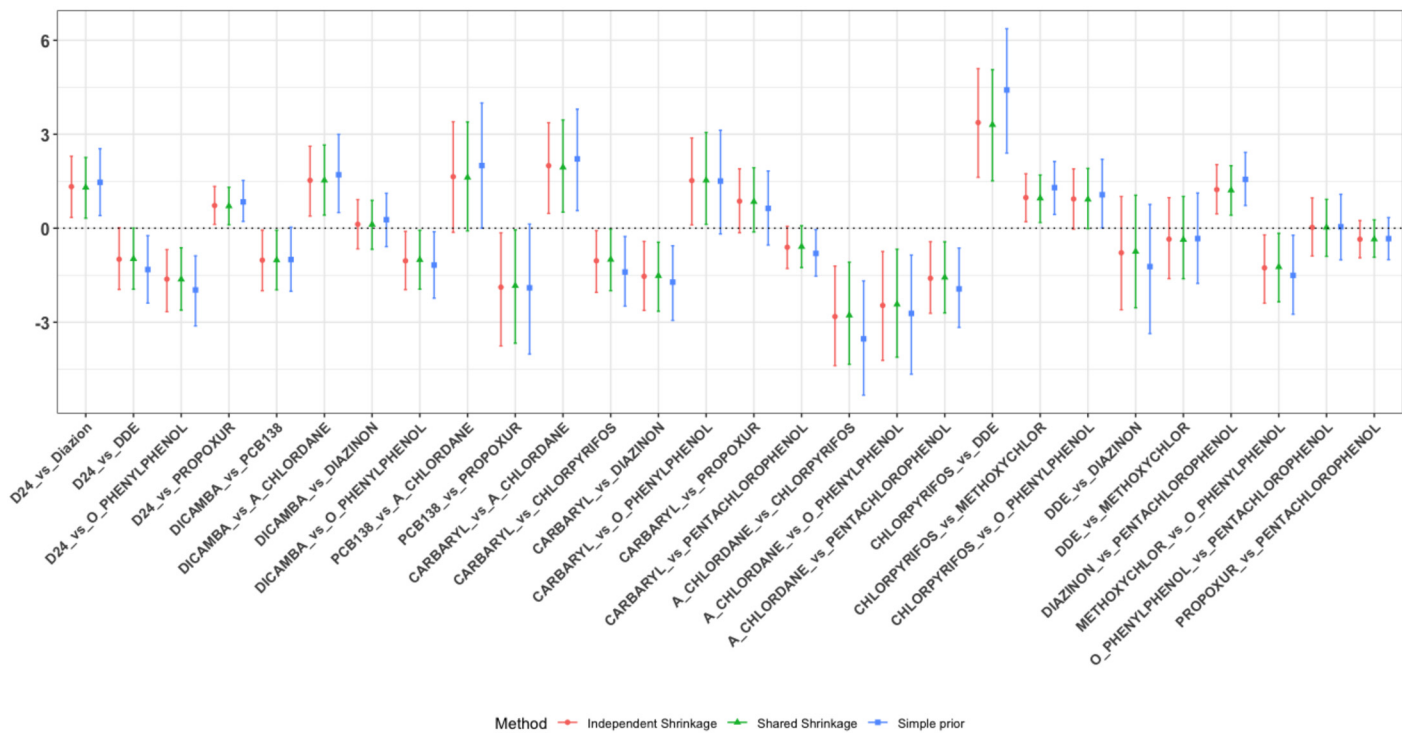

Figure 6: Comparisons between randomly chosen slope vs. slope (γ_{jk4}) interaction effects.

these interactions were with regard to D24. For example, we showed a very strong positive interactive effect between Diazinon and D24 (see Figure 7) that was missed with the linear exposure with the imputation approach.

In the simulations, we considered the case where there were 10 chemicals ($p = 10$); the example contained 14 chemicals in the final models. The analysis results were insensitive to the chosen prior distributions. We used a Gamma(1, 1)
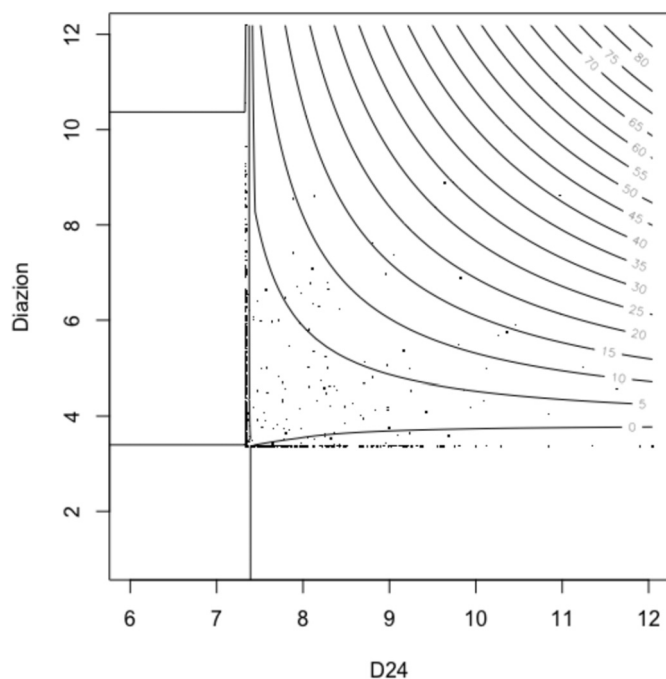
Figure 7: D24 vs. Diazion Contour Plot.

for the local and global shrinkage priors, but results were very similar when we used either Gamma$(1, 0.5)$ with variance 2 or Gamma$(1, 2)$ with variance 0.5 distributions. In the simulation studies, with 10 chemicals we have a large number of interactive effects (45 for linear exposure and 180 for two parameter exposure model). With these large number of interactive effects, the simulations and example showed the importance of using shrinkage priors rather than a vague prior that is closely related to maximum-likelihood. For a larger number of chemicals, even the shrinkage prior approaches may have poor performance due to sparsity. In this case, a Bayesian variable selection method [17] could be used to a smaller number of chemical exposures to include in the model.

We considered models with two-way interactions. Conceptually, we could consider 3 or high-order interactions and extend the shared shrinkage idea to this more complex situation. In the spirit of Stone's generalized additive model [19], the proposed hierarchical model could be extended to include B-splines rather than polynomial effects. Such an extension would be computationally intensive and is left for future work.

DATA AVAILABILITY STATEMENT

Data will be available upon request and data sharing agreement. For R code, see the github page.

REFERENCES

[1] Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**(422) 669–679. MR1224394

[2] Bien, J., Taylor, J. and Tibshirani, R. (2013). A lasso for hierarchical interactions. *Annals of Statistics* **41**(3) 1111. https://doi.org/10.1214/13-AOS1096. MR3113805

[3] Bobb, J. F., Valeri, L., Claus Henn, B., Christiani, D. C., Wright, R. O., Mazumdar, M., Godleski, J. J. and Coull, B. A. (2015). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* **16**(3) 493–508. https://doi.org/10.1093/biostatistics/kxu058. MR3365442

[4] Carrico, C., Gennings, C., Wheeler, D. C. and Factor-Litvak, P. (2015). Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. *Journal of Agricultural, Biological, and Environmental Statistics* **20**(1) 100–120. https://doi.org/10.1007/s13253-014-0180-3. MR3334469

[5] Casella, G., Ghosh, M., Gill, J. and Kyung, M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis* **5**(2) 369–411. https://doi.org/10.1214/10-BA607. MR2719657

[6] Chiou, S. H., Betensky, R. A. and Balasubramanian, R. (2019). The missing indicator approach for censored covariates subject to limit of detection in logistic regression models. *Annals of Epidemiology* **38** 57–64.

[7] Chipman, H., Hamada, M. and Wu, C. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics* **39** 372–381.

[8] Colt, J. S., Lubin, J., Camann, D., Davis, S., Cerhan, J., Severson, R. K., Cozen, W. and Hartge, P. (2004). Comparison of pesticide levels in carpet dust and self-reported pest treatment practices in four US sites. *Journal of Exposure Science and Environmental Epidemiology* **14**(1) 74–83.

[9] Griffin, J., Brown, P. et al. (2017). Hierarchical shrinkage priors for regression models. *Bayesian Analysis* **12**(1) 135–159. https://doi.org/10.1214/15-BA990. MR3597570

[10] Helsel, D. R. (1990). Less than obvious-statistical treatment of data below the detection limit. *Environmental Science & Technology* **24**(12) 1766–1774.

[11] Herring, A. H. (2010). Nonparametric Bayes shrinkage for assessing exposures to mixtures subject to limits of detection. *Epidemiology (Cambridge, Mass.)* **21**(Suppl 4) 71.

[12] Hwang, B. S., Chen, Z., M. Buck Louis, G. and Albert, P. S. (2019). A Bayesian multi-dimensional couple-based latent risk model with an application to infertility. *Biometrics* **75**(1) 315–325. https://doi.org/10.1111/biom.12972. MR3953732

[13] Lange, K. L., Little, R. J. and Taylor, J. M. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**(408) 881–896. MR1134486

[14] Liu, C. (2004). Robit regression: a simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Casual Inference from Incomplete-Data Perspectives* 227–238. https://doi.org/10.1002/0470090456.ch21. MR2138259

[15] McCullagh, P. and Nelder, J. A. (1989). Generalized linear models (2nd edition). Pp 511. £30. 1989. ISBN 0-412-31760-5 (Chapman and Hall). *The Mathematical Gazette* **74**(469) 320–321. https://doi.org/10.2307/3619865.

[16] Ortega-Villa, A. M., Liu, D., Ward, M. H. and Albert, P. S. (2021). New insights into modeling exposure measurements below the limit of detection. *Environmental Epidemiology* **5**(1).

[17] Polson, N. G. and Scott, J. (2010). Shrinkg globally, act locally: Sparse Bayesian regularization and prediction. In *Bayesian Statistics* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) **9** 501–538. https://doi.org/10.1093/acprof:oso/9780199694587.003.0017. MR3204017

[18] Scosyrev, E., Messing, J., Noyes, K., Veazie, P. and Messing, E. (2012). Surveillance Epidemiology and End Results (SEER) program and population-based research in urologic oncology: an overview. In *Urologic Oncology: Seminars and Original Investigations* **30** 126–132. Elsevier.

[19] Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics* 590–606. https://doi.org/10.1214/aos/1176349940. MR0840516

[20] Zahm, S. H., Ward, M. H. and Blair, A. (1997). Pesticides and cancer. *Occupational Medicine (Philadelphia, PA)* **12**(2) 269–289.

[21] Zhang, B., Chen, Z. and Albert, P. S. (2012). Latent class models for joint analysis of disease prevalence and high-dimensional semicontinuous biomarker data. *Biostatistics* **13**(1) 74–88.

Debamita Kundu. Division of Biostatistics, Department of Public Health Sciences, University of Virginia, USA. E-mail address: debamita.kundu@virginia.edu

Sungduk Kim. Biostatistics Branch, Division of Cancer Epidemiology and Genetics National Institutes of Health, USA. E-mail address: kims2@mail.nih.gov

Paul S. Albert. Biostatistics Branch, Division of Cancer Epidemiology and Genetics National Institutes of Health, USA. E-mail address: albertp@mail.nih.gov