



Published in final edited form as:

Biometrics. 2019 March ; 75(1): 315–325. doi:10.1111/biom.12972.

A Bayesian multi-dimensional couple-based latent risk model with an application to infertility

Beom Seuk Hwang¹, Zhen Chen², Germaine M. Buck Louis², Paul S. Albert³

¹Department of Applied Statistics, Chung-Ang University, Seoul, Korea ²Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland ³Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

Abstract

Motivated by the Longitudinal Investigation of Fertility and the Environment (LIFE) Study that investigated the association between exposure to a large number of environmental pollutants and human reproductive outcomes, we propose a joint latent risk class modeling framework with an interaction between female and male partners of a couple. This formulation introduces a dependence structure between the chemical patterns within a couple and between the chemical patterns and the risk of infertility. The specification of an interaction enables the interplay between the female and male's chemical patterns on the risk of infertility in a parsimonious way. We took a Bayesian perspective to inference and used Markov chain Monte Carlo algorithms to obtain posterior estimates of model parameters. We conducted simulations to examine the performance of the estimation approach. Using the LIFE Study dataset, we found that in addition to the effect of PCB exposures on females, the male partners' PCB exposures play an important role in determining risk of infertility. Further, this risk is subadditive in the sense that there is likely a ceiling effect which limits the probability of infertility when both partners of the couple are at high risk.

Keywords

chemical mixture models; couple-based design; latent class model; low dose additivity; subadditivity effect

1 | INTRODUCTION

In many clinical and epidemiological studies, there is interest in identifying patterns of high-dimensional biomarkers and their association with risk of a subsequent disorder. Consider, for example, the Longitudinal Investigation of Fertility and the Environment (LIFE) Study

Correspondence Zhen Chen, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, MD zhen.chen@nih.gov.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2, 4, 5 and 6, as well as R code to implement the method are available with this paper at the Biometrics website on Wiley Online Library.

that enrolled couples trying to become pregnant and followed each couple until pregnant or for 12 months (Buck Louis et al., 2011). A primary interest of the LIFE Study is the relation between mixtures of environmental pollutants (eg, polychlorinated biphenyls (PCBs)) and reproductive endpoints such as infertility, which is defined as 12 months of trying without pregnancy. Complex patterns in the level of these PCB congeners (36 in the LIFE Study) may characterize the risk of infertility.

A unique feature of the LIFE Study lies in its couple-based design that collected environmental exposure data on both partners to study the association between pollutants and infertility. Considering both partners of the couple, the association between chemical exposures and infertility can be complicated since there may be complex interactions between the exposure patterns for each of the two partners. For example, the risk of infertility could be enhanced if both partners of the couple have a certain exposure pattern (ie, synergistic effect). Alternatively, we could have subadditivity where there is little additional effect of one partner's exposure pattern in the presence of a certain pattern for the other partner. Characterizing this interaction is important for understanding the complex environmental effects in couple-based risk assessment.

A naive approach for assessing the interaction effect of chemical exposures between partners and the risk of infertility is to fit, on a chemical by chemical basis, a logistic regression with additive and interaction terms for single chemical exposures on the female and male partners of the couple. Such an approach suffers from the multiplicity problem and does not properly handle the complex patterns in these chemical mixtures. Statistical methodology that examines interactions between patterns of biomarkers is necessary for couple-based risk assessment in this setting.

We propose a Bayesian framework that examines the chemical congener profile from each member of the couple and the risk of infertility. In this work, we link together the complex chemical mixtures on each couple and infertility risk through unobserved latent classes. Specifically, we posit that two sets of latent classes, each characterizing the chemical mixture patterns of one partner of the couple, are linked to the risk of infertility through a logistic model with main and interaction effects between latent classes. This approach allows one to investigate the collective effect of chemical exposures of couples on infertility while avoids the pitfalls of traditional approaches, and represents an advancement in couple-based risk modeling with chemical mixtures.

Joint models using latent class models have been proposed in the literature. Lin et al. (2002) generalized the latent class joint models to analyze longitudinal biomarker and event process data semiparametrically. Elliott et al. (2005) developed a Bayesian latent growth curve model to identify trajectories of positive affect and negative events. Neelon et al. (2011) used Bayesian latent class models to characterize the effect of parity on mental health use and expenditures. Liu et al. (2015) considered latent classes when jointly modeling survival and longitudinal data. Zhang et al. (2012) proposed a latent class model for disease prevalence and chemical mixture data. To our knowledge, our work is the first to focus on using latent class models to link couples' chemical mixture data to disease risk, such as infertility.

We present the proposed modeling framework in Section 2, and specify priors and posterior computations in Section 3. We also conduct model comparison by using the deviance information criterion (DIC) in Section 3. In Section 4, we apply our approach to the LIFE Study data, and conduct simulation studies to demonstrate the performance of the proposed approach in Section 5. We conclude with a discussion of our approach and directions for future work in Section 6.

2 | METHODS

2.1 | Joint latent class model

The motivation of the proposed joint latent class model concerns a study dataset where a hierarchical structure exists, with a couple consisting of male and female partners and concentrations of a collection of PCB congeners measured for each of the partners. Let Y_i be a binary variable indicating fertility or infertility for the i th couple, $i = 1, \dots, I$, where $Y_i = 1$ denotes infertility and $Y_i = 0$ denotes fertility. Let X_{ij}^F and X_{ij}^M be the concentrations of the j th PCB exposure measured in serum for female and male partners, respectively in the i th couple, $j = 1, \dots, J$. We introduce latent class variables for females and males, L_i^F, L_i^M to account for the risk of infertility, where $L_i^F (L_i^M)$ takes the value $k (k = 0, \dots, K - 1)$ if female (male) in the i th couple belongs to class k . Although we use the same K for both partners here for presentational brevity, we will allow partner specific K s in actual implementations of the methods in Sections 4 and 5. We assume the latent class has higher risk of infertility as k increases; for example, the 3-latent class model is composed of low-risk class ($L_i = 0$), medium-risk class ($L_i = 1$) and high-risk class ($L_i = 2$). Let $\pi_k^F = Pr(L_i^F = k)$ and $\pi_k^M = Pr(L_i^M = k)$ be the probability of a male and a female, respectively belonging to latent class k . The probability distribution of infertility of couple i given the latent class variables is expressed as:

$$\text{logit}P(Y_i = 1 | L_i^F, L_i^M) = \beta_0 + \beta_1^F L_i^F + \beta_1^M L_i^M + \beta_2 L_i^F L_i^M, \quad (1)$$

where $\beta_0, \beta_1^F > 0, \beta_1^M > 0$, and β_2 are regression coefficients representing the association between the risk of infertility and the latent class variables. The coefficient β_0 represents the log odds of infertility when both partners are in the lowest risk classes. β_1^F is interpreted as the change of the log odds of infertility from a female's risk class to the next higher risk class given the corresponding male belongs to the lowest risk class; β_1^M has similar interpretations. The interaction coefficient β_2 can be interpreted as how differences in log odds between two adjacent females' latent classes are different depending on what risk classes the corresponding males belong to. Depending on the sign, the interaction effect implies synergistic effect or subadditivity effect of risk classes between females and males.

In the LIFE Study data, about 24% of PCB values are too low to detect (hence are zeros). Such semicontinuous data need to be modeled through a mixture of a degenerate distribution at zero and a continuous distribution for nonzero values (Manning et al., 1981; Cooper et al., 2003; Su et al., 2009; Liu et al., 2010, 2012; Neelon et al., 2011). This 2-part model

incorporates distinctive mechanisms for the zero and non-zero parts. Alternatively, a Tobit modeling framework can be used if the zeros are regarded as censored values. The nonzero parts seem to have right-skewed distributions (see Figure 1 for congener 153). These distributions of all PCB congeners are much less skewed on logarithm scale (see Figure S1 in Supplementary Materials). Based on these features, each X_{ij} can be represented by two variables: for $i = 1, \dots, I, j = 1, \dots, J, G \in \{F, M\}$,

$$U_{ij}^G = \begin{cases} 1, & \text{if } X_{ij}^G \neq 0 \\ 0, & \text{if } X_{ij}^G = 0 \end{cases} \text{ and } V_{ij}^G = \begin{cases} X_{ij}^G, & \text{if } X_{ij}^G \neq 0 \\ \text{irrelevant}, & \text{if } X_{ij}^G = 0 \end{cases} \quad (2)$$

where U_{ij}^G is the binary nonzero PCB value indicator for partner G and V_{ij}^G the nonzero value of the PCB exposures with log-normal distributions (see Figure 1c and d)

$$V_{ij}^G | L_{ij}^G, \mathbf{b}_j^G \sim \log N(\mu_{ij}^G(L_i^G, \mathbf{b}_j^G), \tau_G^2), \quad (3)$$

with $\mu_{ij}^G(L_i^G, \mathbf{b}_j^G)$ and τ_G^2 denoting the mean and variance of V_{ij}^G on the log scale. The log-normality assumption ensures that the V_{ij}^G 's have positive support. We assume that $\mu_{ij}^G(L_i^G, \mathbf{b}_j^G)$ depends on the latent class variable L_i^G and random effects vector \mathbf{b}_j^G . The latent class variables allow the PCBs from the same participant to be correlated, and the random effects allow for each PCB to have varying departures from the overall mean:

$$\mu_{ij}^G(L_i^G, \mathbf{b}_j^G) = \alpha_0^G + \alpha_1^G L_i^G + b_{0j}^G + b_{1j}^G L_i^G, \quad (4)$$

where α_0^G and α_1^G are fixed effects coefficients of the latent class variables, and \mathbf{b}_j is the PCB-specific shared random effects vector across females and males following a multivariate normal distribution, $\mathbf{b}_j = (b_{0j}^F, b_{1j}^F, b_{0j}^M, b_{1j}^M)' \sim N_4(\mathbf{0}, \Sigma_b)$, where

$$\Sigma_b = \begin{pmatrix} \sigma_{0F}^2 & \rho_{01}^{FF} & \sigma_{0F}\sigma_{1M} & \rho_{00}^{FM} & \sigma_{0F}\sigma_{0M} & \rho_{01}^{FM} & \sigma_{0F}\sigma_{1M} \\ & \sigma_{1F}^2 & \rho_{10}^{FM} & \sigma_{1F}\sigma_{0M} & \rho_{11}^{FM} & \sigma_{1F}\sigma_{1M} \\ & & \sigma_{0M}^2 & \rho_{01}^{MM} & \sigma_{0M}\sigma_{1M} \\ & & & \sigma_{1M}^2 \end{pmatrix},$$

and σ_{0G}^2 is the variance of intercepts of random effects, σ_{1G}^2 is the variance of slopes of random effects, and $\rho_{ll'}^{G_1 G_2}$ is the correlation between l for gender G_1 and l' for gender G_2 , $l, l' = 0$ (intercepts) or 1 (slopes), and $G, G_1, G_2 = \text{females or males}$.

To achieve model parsimony, we assume that the probability of nonzero PCB is associated with the mean of nonzero PCB values $\mu_{ij}^G(L_i^G, \mathbf{b}_j^G)$ as follows:

$$\text{logit}P(U_{ij}^G = 1 \mid L_i^G, \mathbf{b}_j^G) = \eta_0^G + \eta_1^G h(\mu_{ij}^G(L_i^G, \mathbf{b}_j^G)), \quad (5)$$

where η_0^G and η_1^G are coefficients, and function $h(\cdot)$ is specified based on data structure.

Figure S2 in Supplementary Materials suggests linear relationship between the probability of nonzero value and the mean of nonzero values for both females and males, that is, $h(\mu_{ij}^F(L_i^F, \mathbf{b}_j^F)) = \mu_{ij}^F(L_i^F, \mathbf{b}_j^F)$ and $h(\mu_{ij}^M(L_i^M, \mathbf{b}_j^M)) = \mu_{ij}^M(L_i^M, \mathbf{b}_j^M)$. The chemical concentration model (Equations (2)–(5)) is linked with the infertility model (Equation (1)) through the latent class variables. The proposed model can be reduced to simpler ones by placing additional constraints on the β 's in Equation (1). When $\beta_1^F \neq 0, \beta_1^M \neq 0$ and $\beta_2 = 0$ and $\beta_2 = 0$, both female and male of the couple independently have effects on the risk of infertility, but there are no interaction effects within couple. When there are no correlations between the female and male's PCBs ($\rho_{00}^{FM} = \rho_{01}^{FM} = \rho_{10}^{FM} = \rho_{11}^{FM} = 0$), $\beta_2 = 0$, $\beta_2 = 0$ and either $\beta_1^F = 0$ or $\beta_1^M = 0$, the model becomes a single gender model as in Zhang et al. (2012). Given the data $\{y_i, x_{ij}^F, x_{ij}^M\}$ or equivalently $\{y_i, u_{ij}^F, v_{ij}^F, u_{ij}^M, v_{ij}^M\}$, the complete data likelihood is given by

$$\begin{aligned} L = & \prod_{i=1}^I \left(\frac{e^{\Lambda_i}}{1 + e^{\Lambda_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\Lambda_i}} \right)^{1-y_i} \pi_{L_i^F}^F \pi_{L_i^M}^M \\ & \times \prod_{i=1}^I \prod_{j=1}^J \left(\frac{e^{\eta_0^F + \eta_1^F \mu_{ij}^F(\cdot)}}{1 + e^{\eta_0^F + \eta_1^F \mu_{ij}^F(\cdot)}} \right)^{u_{ij}^F} \left(\frac{1}{1 + e^{\eta_0^F + \eta_1^F \mu_{ij}^F(\cdot)}} \right)^{1-u_{ij}^F} \\ & \times \prod_{i=1}^I \prod_{j=1}^J \left(\frac{e^{\eta_0^M + \eta_1^M \mu_{ij}^M(\cdot)}}{1 + e^{\eta_0^M + \eta_1^M \mu_{ij}^M(\cdot)}} \right)^{u_{ij}^M} \left(\frac{1}{1 + e^{\eta_0^M + \eta_1^M \mu_{ij}^M(\cdot)}} \right)^{1-u_{ij}^M} \\ & \times \prod_{i=1}^I \prod_{j=1}^J \left[\log N(v_{ij}^F; \mu_{ij}^F(L_i^F, \mathbf{b}_j^F), \tau_F^2) \right]^{u_{ij}^F} \\ & \times \left[\log N(v_{ij}^M; \mu_{ij}^M(L_i^M, \mathbf{b}_j^M), \tau_M^2) \right]^{u_{ij}^M} \times \prod_{j=1}^J N_4(\mathbf{b}_j; \mathbf{0}, \Sigma_b), \end{aligned} \quad (6)$$

where $\Lambda_i = \beta_0 + \beta_1^F L_i^F + \beta_1^M L_i^M + \beta_2 L_i^F L_i^M$.

2.2 | Covariates dependence

The proposed joint latent class model can be extended to accommodate covariates adjustment. Subject-specific covariates can be incorporated into the distribution of the nonzero PCB values (3) and the model of the nonzero PCB indicator (5) through the conditional mean of the nonzero PCB exposures (4) as follows:

$$V_{ij}^G \mid L_{ij}^G, \mathbf{b}_j^G, \mathbf{W}_i^G \sim \log N(\mu_{ij}^G(L_i^G, \mathbf{b}_j^G, \mathbf{W}_i^G), \tau_G^2), \quad (7)$$

$$\text{logit}P(U_{ij}^G = 1 \mid L_i^G, \mathbf{b}_j^G, \mathbf{W}_i^G) = \eta_0^G + \eta_1^G h(\mu_{ij}^G(L_i^G, \mathbf{b}_j^G, \mathbf{W}_i^G)), \quad (8)$$

$$\mu_{ij}^G(L_i^G, \mathbf{b}_j^G, \mathbf{W}_i^G) = \alpha_0^G + \alpha_1^G L_i^G + b_{0j}^G + b_{1j}^G L_i^G + \mathbf{W}_i^{G'} \lambda^G, \quad (9)$$

where \mathbf{W}_i^G is the vector of subject-specific covariates such as age, BMI or smoking status, and λ^G is parameter vector. We note that covariates can also be included in the infertility model (1) and potentially in class membership component $\pi_k^G = \Pr(L_i^G = k)$. However, care must be taken in incorporating covariates in class membership, as the latent class variables L_i^G 's depend on covariates through the other two model components.

3 | PRIORS AND POSTERIOR COMPUTATIONS

3.1 | Prior specification

We assume independent prior distributions for $\beta_0, \beta_1^F, \beta_1^M, \beta_2$ in the infertility model (Equation (1)), in which $\beta_0 \sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2)$, $\beta_1^F \sim \log N(\mu_{\beta_1^F}, \sigma_{\beta_1^F}^2)$, $\beta_1^M \sim \log N(\mu_{\beta_1^M}, \sigma_{\beta_1^M}^2)$ and $\beta_2 \sim N(\mu_{\beta_2}, \sigma_{\beta_2}^2)$. The log-normal prior distributions for β_1^F and β_1^M were used to reflect positive coefficient assumption. We assign weakly informative prior hyperparameters for log-normal distributions to prevent the posterior distributions from moving away too extremely in MCMC, while we assign noninformative prior hyperparameters for the normal distribution for β_0 and β_2 .

In the PCB exposure model, we first assign independent noninformative normal prior distributions for $\alpha_0^F, \alpha_1^F, \alpha_0^M$, and α_1^M : $(\alpha_0^F, \alpha_1^F, \alpha_0^M, \alpha_1^M)' \sim MVN(\mu_\alpha, \Sigma_\alpha)$. The random effect covariance matrix, Σ_b is assumed to have a conjugate inverse-Wishart distribution: $\Sigma_b \sim IW(v_0, \Sigma_0)$. We assume that the scale parameters of log-normal distributions for positive PCB exposures have conjugate inverse-Gamma distributions: $\tau_{F^2}^2 \sim IG(a_{\tau_F}, b_{\tau_F})$, $\tau_{M^2}^2 \sim IG(a_{\tau_M}, b_{\tau_M})$. Further, we take independent noninformative normal prior distributions for $\eta_0^F, \eta_1^F, \eta_0^M$, and η_1^M , and also for λ^F and λ^M . The class-membership probabilities, π_k^F, π_k^M , have conjugate Dirichlet prior distributions: $(\pi_1^F, \dots, \pi_K^F) \sim \text{Dirichlet}(e_1^F, \dots, e_K^F)$, $(\pi_1^M, \dots, \pi_K^M) \sim \text{Dirichlet}(e_1^M, \dots, e_K^M)$. Finally, we assume that the number of latent classes, K is known and is allowed to differ in females and males. We choose K using model comparison technique as discussed in Section 3.3.

3.2 | Posterior computations

Posterior computation proceeds with the Metropolis-Hastings within Gibbs algorithm. In the mixed effects model of the mean structure of positive PCB exposures (Equation (4)), we observed strong correlation between the conditional posterior distributions of fixed effects and random effects (weak identifiability). This often occurs with a noninformative prior specification for model parameters and results in slow convergence and bad mixing in

MCMC procedures. To address this issue, we consider the hierarchical centering reparametrization (Gelfand et al., 1995) that can result in less correlated parameters *a posteriori*. In the case of no covariates (Equation (4)), we redefine the mean structure by moving the fixed effects to the mean of random effects:

$$\begin{aligned}\mu_{ij}^G(L_i^G, \mathbf{b}_j^G) &= (\alpha_0^G + b_{0j}^G) + (\alpha_1^G + b_{1j}^G)L_i^G \\ &= b_{0j}^{G*} + b_{1j}^{G*} L_i^G, G \in \{F, M\}, \mathbf{b}_j \\ &= (b_{0j}^{F*}, b_{1j}^{F*}, b_{0j}^{M*}, b_{1j}^{M*})' \\ &\sim N_4((\alpha_0^F, \alpha_1^F, \alpha_0^M, \alpha_1^M)', \Sigma_b).\end{aligned}\tag{10}$$

A similar reparameterization is made when covariates are added. This reparametrization results in improved behavior in the MCMC algorithm. Further, in the probability distribution of nonzero PCB exposure (Equation (5)), another centering method was used to improve the MCMC algorithm. Equation (5) can be rewritten as:

$$\begin{aligned}\text{logit}P(U_{ij}^G = 1 \mid L_i^G, \mathbf{b}_j^G) &= \eta_0^G + \eta_1^G \mu_{ij}^G(\cdot) = \eta_0^G + \eta_1^G (\mu_{ij}^G(\cdot) - \bar{\mu}^G) \\ &+ \eta_1^G \bar{\mu}^G = (\eta_0^G + \eta_1^G \bar{\mu}^G) + \eta_1^G (\mu_{ij}^G(\cdot) - \bar{\mu}^G) \\ &= \eta_0^{G*} + \eta_1^G (\mu_{ij}^G(\cdot) - \bar{\mu}^G),\end{aligned}\tag{11}$$

The parameters, η_0^{F*} and η_0^{M*} are estimated by MCMC instead of η_0^F and η_0^M , resulting in much better mixings. In the case where the full conditional distributions do not have closed forms, we use an adaptive Metropolis algorithm (Haario et al., 2005) to update the parameters. The proposal distribution is updated with a normal distribution with the empirical covariance of the chain sampled up to that point. The covariance is then adjusted by the scaling factor to achieve an optimal acceptance rate of approximately 0.44 (Gelman et al., 2014). A complete description and derivation of the full conditional posterior distributions and the MCMC updating steps are provided in Sections S1 and S2 in Supplementary Materials. All analyses were conducted in R version 3.2.2 on the NIH Biowulf cluster.

3.3 | Model comparison

Several model selection criteria can be used to select the number of latent classes from a Bayesian perspective, including the deviance information criterion (DIC) developed by Spiegelhalter et al. (2002). The DIC is defined as $DIC = \overline{D(\theta)} + p_D$, where $D(\theta) = -2\log f(y \mid \theta)$ is the deviance, $\overline{D(\theta)}$ is the posterior mean deviance and $p_D = \overline{D(\theta)} - D(\tilde{\theta})$ is the effective number of parameters in a model. The DIC considers both a measure of fit (deviance) and a measure of complexity (p_D). However, p_D may be negative for models that do not have a log-concave density, that is, in the cases of mixtures of distributions and random effect models (Celeux et al., 2006). As a result, we use a modified DIC proposed by Celeux et al as follows:

$$\text{DIC}_4 = -4E_{\theta, Z}[\log f(y, Z | \theta) | y] \\ + 2E_Z[\log f(y, Z | E_{\theta}[\theta | y, Z]) | y]$$

where $f(y, Z | \theta)$ is the complete likelihood, y is the observed data and Z are the random effects and latent variables. We note that the Watanabe-Akaike information criterion (WAIC) proposed by Watanabe (2010) is gaining popularity in Bayesian literature and can be potentially used.

4 | LIFE DATA ANALYSIS

4.1 | Data descriptions

In the LIFE Study, 501 couples were enrolled, of which 401 (80%) were followed until an hCG (human chorionic gonadotropin) pregnancy or 12 months of trying without pregnancy. As our analysis is concerned with infertility (defined as a prospectively observed time to pregnancy greater than or equal to 12 months, consistent with the clinical definition of infertility (Practice Committee of the American Society for Reproductive Medicine, 2013)), these 401 couples form our analytical working data. Among them, 347 couples became pregnant while 54 did not during the study period. A detailed description of the LIFE Study design and main study outcomes can be found elsewhere (Buck Louis et al., 2011).

Concentrations of 36 PCB congeners were obtained for both partners of the couple at baseline. While most of these concentrations are positive, a considerable portion (24.2% overall) are true measured zeros. A small percentage (<0.5%) of the measured PCB concentrations are negative, as a result of the laboratory quality control process. We set these concentrations to missing in the analysis. However, no participant is deleted from the analysis since none is completely missing PCB data. The distributions of these PCB congeners are presented in Figure S1 in Supplementary Materials.

Several partner-specific covariates were identified *a priori* to be potential confounders for the PCB concentrations and infertility, including age (*years*; continuous), BMI (kg/m^2 ; categorized into <25, 25–29, 30), serum lipids (ng/g ; continuous) and smoking status (serum cotinine level (ng/mL); continuous). Serum lipids were log-transformed and serum cotinine log-transformed after one was added to avoid taking the logarithm of negative values. We adjusted for serum lipids because PCBs are lipophilic chemicals. Age, lipids, and cotinine were all standardized to simplify computations prior to analysis. The distributions of the covariates on the original scale are reported separately in Table 1 for female and male partners of the couple.

We created four couple-level variables for potential inclusion in the infertility model Equation (1): female age, difference between male and female ages, average BMI, and average serum cotinine level. Since parameter estimates are quite similar with and without these covariates (see Table S1 in Supplementary Materials), we proceed with the analysis without further considering covariates in the infertility model.

4.2 | Model specifications

The primary objective of this analysis is to investigate the association between multi-dimensional PCB exposures and couple's infertility risk linked through underlying latent classes. To fit the proposed model, we take the following values of hyperparameters in the prior distributions.

$$\begin{aligned}
 \beta_0 &\sim N(0, 100), & \beta_1^F &\sim \log N(0, 1) \\
 \beta_1^M &\sim \log N(0, 1), & \beta_2 &\sim N(0, 100) \\
 \alpha_0^F &\sim N(0, 100), & \alpha_1^F &\sim N(0, 100) \\
 \alpha_0^M &\sim N(0, 100), & \alpha_1^M &\sim N(0, 100) \\
 \eta_0^F &\sim N(0, 100), & \eta_1^F &\sim N(0, 100) \\
 \eta_0^M &\sim N(0, 100), & \eta_1^M &\sim N(0, 100) \\
 (\pi_1^F, \dots, \pi_K^F) &\sim \text{Dirichlet}(1, \dots, 1) \\
 (\pi_1^M, \dots, \pi_K^M) &\sim \text{Dirichlet}(1, \dots, 1) \\
 \tau_F^2 &\sim IG(1, 1), & \tau_M^2 &\sim IG(1, 1), & \Sigma_b &\sim IW(4, I_4), \\
 \lambda^F &\sim MVN(\mathbf{0}_5, 100I_5), & \lambda^M &\sim MVN(\mathbf{0}_5, 100I_5).
 \end{aligned}$$

These values were chosen to induce proper but vague priors. We obtained results by running the MCMC algorithm for 100 000 iterations with 50 000 iteration burn-ins. The mixing of the MCMC iterations are satisfactory in both the fixed effects and variance components, as demonstrated by the trace plots in Supplementary Materials (see Figures S3 and S4 in Supplementary Materials).

4.3 | Results

To establish the best model, we let the number of risk classes vary between 2 and 8 for both partners. The model selection criteria DICs are presented in Table 2. While we saw that the model with seven risk classes for both partners provides the lowest DIC, we also observed that the prevalence of classes higher than 5 is extremely small and that the posterior estimates of the model parameters are very close to each other and to those from the 5-class model. Thus, for model parsimony, our interpretations are based on the 5-class model.

Tables 3 provides parameter estimates of the best model (five classes for both partners) in addition to the results from models where both partners can be grouped into 2-, 3-, or 4-classes. To save space, we did not include estimates from other models. Overall, the estimates are similar, especially after the 3-class model. Figure 2 also shows that the prevalence of the highest risk class decreases rapidly to zero after the 3-class

model, suggesting the results have stabilized. Focusing on the 5-class model in Table 3, we see that the odds of infertility are about 40% ($=\exp(0.34)$) higher when the female partner of the couple moves to a higher risk class, if the male partner is in the lowest risk class ($L_i^M = 0$). The corresponding number for the male partner is slightly higher: the odds of

infertility are about 60% ($=\exp(0.47)$) higher when the male partner of the couple moves to a higher risk class, if the female partner is in the lowest risk group ($L_i^F = 0$). The negative estimate of the interaction effect ($\hat{\beta}_2 = -0.27$) indicates that a couple's risk of infertility does not necessarily increase when one partner moves to a higher risk class, implying a subadditivity effect. Figure S5 and Table S2 in Supplementary Materials provide estimated probabilities of infertility by risk classes and the corresponding credible intervals. This phenomenon of subadditivity has been discussed in the chemical mixture field (Braun et al., 2016). Although the figure suggests that the risks for the 4–4 group is lower than the 0–0 group, the credible intervals are very wide. Thus, we can safely conclude a negative interaction and thus subadditivity, but not draw strong inferences about the estimated probabilities when both latent classes are three or more.

The estimates of α_1^F and α_1^M are positive, which implies that higher-risk classes are more likely to have large mean values of nonzero PCB exposures than lower-risk classes. The estimates of η_1^F and α_1^M are also positive, suggesting that the probability of having positive PCB exposures increases with the mean PCB concentrations. The variance component estimates of the PCB-specific random effects indicate that there are strong positive correlations between female and male partners, in both random intercepts ($\rho_{00}^{FM} = 0.97$) and slopes ($\rho_{11}^{FM} = 0.36\text{--}0.55$). This is not surprising, as the correlation between average female and male PCB concentrations in the data is about 0.989.

Figure S8 in Supplementary Materials shows the estimated means of nonzero PCB concentrations by latent risk class for the best model. Several observations can be made: (1) The patterns in females and males are quite similar; (2) Except in a few congeners (eg, the first four), the estimated means of the concentrations are well separated between risk groups, suggesting that the effect of the PCBs on infertility is quite ubiquitous albeit small. Individually, these low dose effects might not be captured easily; but their collective effect can be estimated through the proposed model.

The proposed modeling approach places strong linearity assumptions in the risk and chemical components of the model. We showed the adequacy of these assumptions on the risk using various analytical strategies (see discussions in Supplementary Materials).

To assess goodness of fit of the modeling framework, we separately examined each of the two modeling components. For the infertility model in Equation (1), we obtained the predicted values of y_i 's and compared them to the observed risk of infertility by deciles. As illustrated in the top panel of Figure 3, the fitted values are close to the observed, suggesting a reasonable fit of the infertility model component. There does seem to be a slightly attenuated pattern for the predicted compared with the observed. When we considered models with three latent classes for both partners, we found a similar “lack of fit” (the bottom two panels of Figure 3). In fact, the 3-class model using the flexible indicator variable framework showed somewhat worse calibration. This lack of fit for both at the extremes is consistent with chance due to the low prevalence of infertility and small sample sizes. For the chemical concentration model as specified in Equations (7)–(9), we compared

the residuals of each PCB congener between fertile and infertile couples, separately for male and female partners. As illustrated in Figures S9 and S10 in Supplementary Materials, the fertile and infertile populations have similar distributions for the residuals, indicating that the model removes any effect of infertility from the PCB profiles. Further analysis (discussed in Supplementary Materials) also showed that latent classes are well discriminated in our application.

While it is possible to have congener-specific variance components, doing so would greatly increase model complexity. As a sensitivity analysis, we have used the empirical LIFE Study dataset to divide the 36 congeners into three groups (by tertiles of empirical variances) and allowed the model to have three different variances. Results under such a new model are similar to the old ones (see Table S4), suggesting that the LIFE Study data analysis results are robust to different specifications of the variance components.

5 | SIMULATION STUDIES

We conducted simulation studies and found the model has good operating characteristics. In particular, we found that the proposed modeling framework is capable of handling equal- as well as unequal-number of risk classes, that the developed MCMC algorithm produces posterior estimates that are very close to the true parameter values, that the DIC always chooses the models with correct number of risk classes in female and male partners, and that the convergence and mixing of the MCMC are satisfactory. Details of the simulation studies are included in Section S3 of Supplementary Materials.

6 | DISCUSSION

We proposed a Bayesian joint latent class model of high-dimensional chemical exposures and the risk of infertility. Exposures to a collection of PCB congeners are linked to the risk of infertility through the latent risk classes of both partners of the couple. The latent class variables allow the risk of infertility to differ across the classes, and differ between the two partners of a couple. The proposed model considers the collective effect of exposures to all PCB congeners, each potentially too small to be significant when examined individually through standard approaches. Further, the model takes into account the correlated exposure patterns of both partners of the couple while considering the complex interactions between them. The proposed use of a joint modeling approach (for couples) to treat low-dose chemical effects provides an additional arsenal in the analytic tool for modeling chemical exposures and health outcomes. Although our model was demonstrated through a human fecundity study, it is clearly applicable to a wide array of applications in obstetric and pediatric medicine where both female and male partners may influence the outcome of their child.

The application of our proposed approach to the LIFE Study yields some insightful empirical findings. The main effects of both partners in the infertility model (β_1^F and β_1^M) are positive with relatively large magnitudes (ORs = 1.4 and 1.6, respectively), suggesting that male PCB exposure needs to be carefully considered in assessing the effect of environmental contaminants on infertility. Interestingly, all our models suggest a negative interaction

between partner-specific latent classes and the risk of infertility. This was true when we fit models with 2–8 latent classes for each partner. The negative interaction suggests that once one partner of the couple has a high risk chemical exposure pattern, then the other partner's risk profile does not increase the risk of infertility.

Our model assumes that all of the correlation between congeners is induced through the latent class. We understand this limitation, but also recognize that a model that incorporated a saturated conditional correlation structure would not be identifiable. As an alternative, we examined the robustness of the model to departure of the conditional dependence structure. To that end, we have conducted simulations to investigate the implications of ignoring conditional dependence between congeners. With 3 latent classes for both genders, we let the PCBs within a class to be correlated at different values, with congeners 1–4, 5–8, 9–12, 13–36 correlated at 0.2, 0.3, 0.4, and 0 levels, respectively. We also generated data with all PCB congeners being uncorrelated within a class. Average parameter estimates based on these two scenarios (correlated and uncorrelated) are shown in Table S5 in Supplementary Materials. We see no substantial differences in these average point estimates between different correlation scenarios. Moreover, all estimates are close to the true values. These observations suggest that the impact of not fully accounting for reasonable conditional dependence across congeners is minimum.

In models with latent classes, convergence can be an issue if the model is unidentified with respect to label switching. We follow the literature (Lenk and DeSarbo, 2000; Congdon, 2005) to impose a simple inequality constraint on the coefficients of latent class variables so that the higher indexed class has higher risk of infertility, that is, $\beta_1^F > 0$ and $\beta_1^M > 0$. In our simulations and real data example, we ran our algorithm with different starting values to make sure the MCMC chain converges to the same distribution. Figure S12 in Supplementary Materials presents the trace plots for β 's with different starting points in Scenario I of one realization of our simulation study. We see that the β 's rapidly converge to the true values from different starting points, indicating no problem of multiple convergence points.

In the chemical concentration model, we have specified random effects across j to allow for the relationship between concentrations and latent severity L_j to vary across PCB congeners. This is a parsimonious way to incorporate the heterogeneous relationship between concentration and severity by PCB congeners. While when marginalized over j there is a dependence among common PCB concentrations on independent subjects, we do not think that the model should be interpreted in this way. Rather the interpretation should be conditional on the particular PCB congener j , and in this case, the concentrations are independent across individuals.

We have taken a Bayesian perspective in conducting inference with the proposed models. An advantage of the Bayesian approach to fitting the latent class model is the efficient MCMC computation. Maximum-likelihood methods have great difficulty in estimating the model parameters since the likelihood function involved high-dimensional integration. The Monte Carlo EM algorithm resulted in bad mixing and poor convergence in this study. The use of hierarchical centering reparameterizations and adaptive Metropolis-Hastings algorithm

improved MCMC computation in this model. The Bayesian approach also has advantages of the incorporation of prior belief, efficient standard error estimation, and not resorting to asymptotic approximations (Gelman et al., 2014).

In some applications, the latent risk class might not be static and can evolve over time. When longitudinal data are available on the chemical exposures, a dynamic modeling framework can be constructed where the latent risk classes can follow a Markov process with distinctive transition probabilities (hidden Markov models). It is also possible to let the latent risk class to depend on subject- or couple-specific covariates. Developing such models is an area of future research.

Other potential extensions of the proposed approach are possible. In the current work, we have assumed that the latent risk classes L_i^F and L_i^M enter the model as ordinal variables mainly for a parsimonious modeling framework. However, it might be better to treat them as discrete if the linearity assumption is deemed inappropriate or to model the classes using parametric functions if it is believed a large number of latent classes exist. Finally, a spike and slab prior can be used for the main effects (β_1^F and β_1^M) in the infertility model. Such a prior will enable us to explicitly test the hypotheses that these coefficients are zeros.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

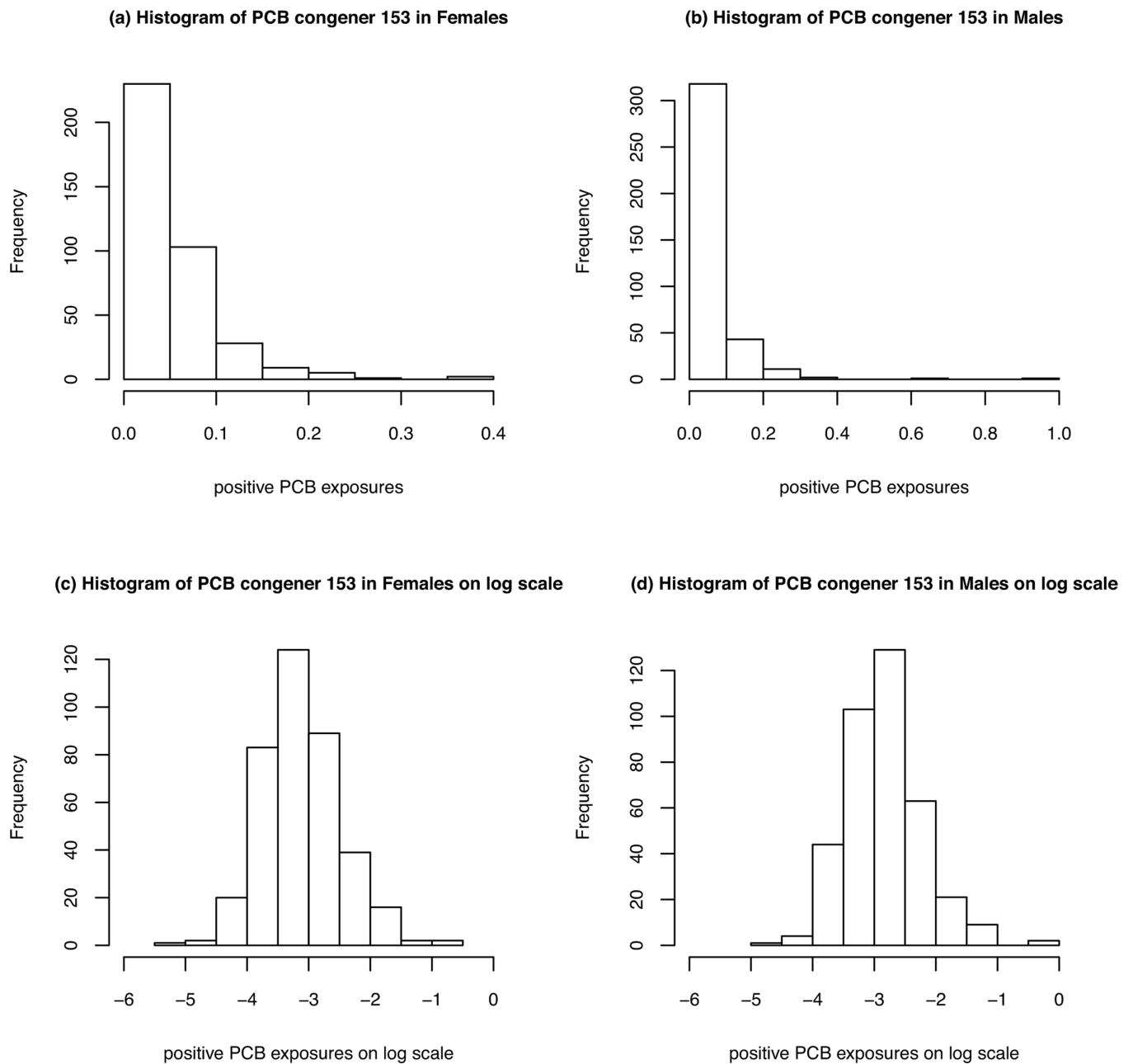
ACKNOWLEDGEMENTS

We thank the co-editor, associate editor, and referees for their helpful comments that considerably improved the article. Chen, Albert, and Buck Louis were supported by the Intramural Research Program of Eunice Kennedy Shriver National Institute of Child Health and Human Development, and Hwang was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03933334). This work utilized the computational resources of the NIH HPC Biowulf cluster.

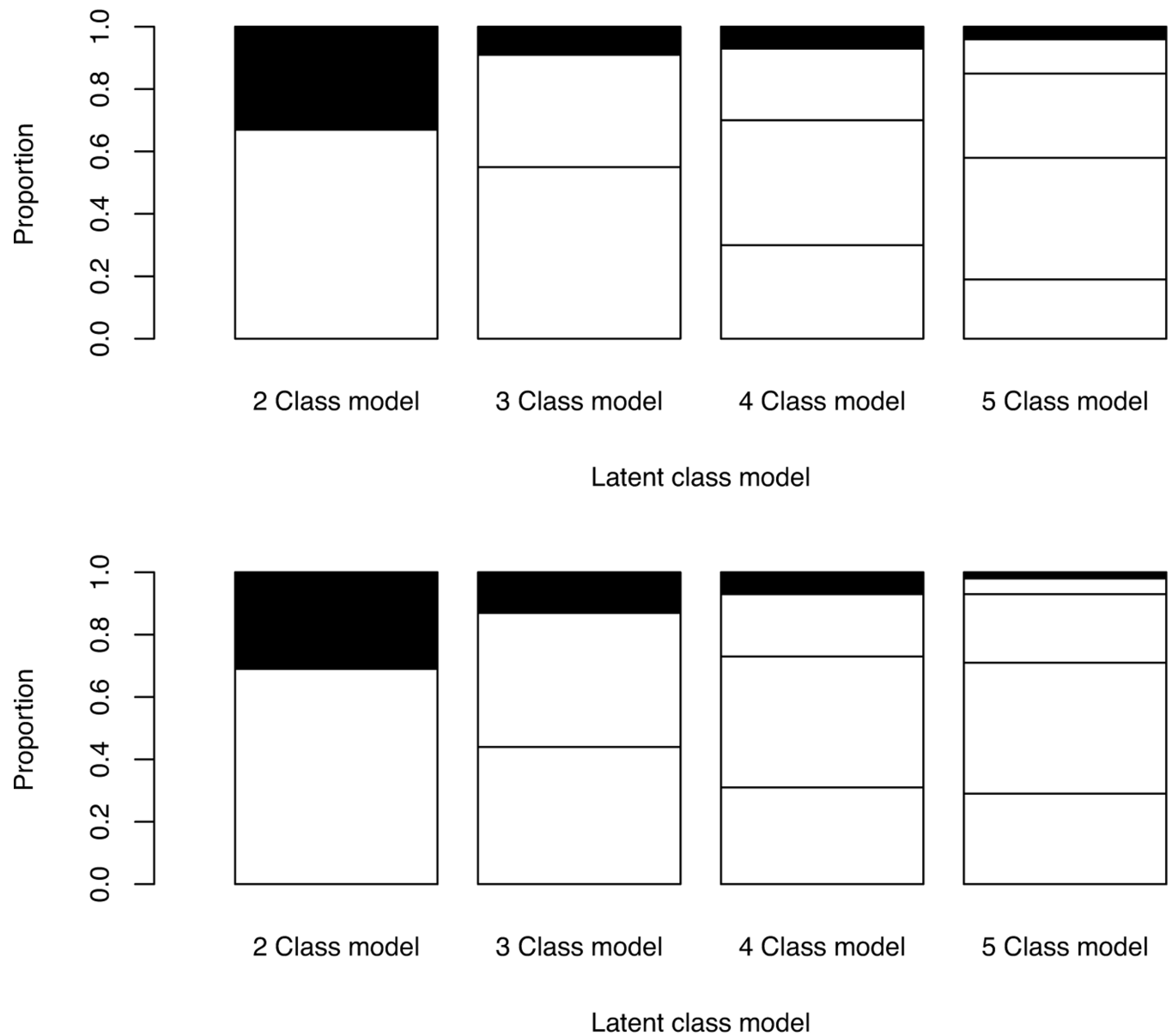
REFERENCES

- Braun JM, Gennings C, Hauser R, and Webster TF (2016). What can epidemiological studies tell us about the impact of chemical mixtures on human health. *Environ Health Perspect* 124, A6–9. [PubMed: 26720830]
- Buck Louis GM, Schisterman EF, Sweeney AM, et al. (2011). Designing prospective cohort studies for assessing reproductive and developmental toxicity during sensitive windows of human reproduction and development—The LIFE Study. *Paediatr Perinat Epidemiol* 25, 413–424. [PubMed: 21819423]
- Celeux G, Forbes F, Robert CP, and Titterton DM (2006). Deviance information criteria for missing data models. *Bayesian Analysis* 1, 651–674.
- Congdon P. (2005). *Bayesian models for categorical data*. Chichester, U.K: John Wiley & Sons.
- Cooper NJ, Sutton AJ, Mugford M, and Abrams KR (2003). Use of Bayesian Markov chain Monte Carlo methods to model cost-of-illness data. *Med Decis Making* 23, 38–53. [PubMed: 12583454]
- Elliott MR, Gallo JJ, Ten Have TR, Bogner HR, and Katz IR (2005). Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. *Biostatistics* 6, 119–143. [PubMed: 15618532]
- Gelfand AE, Sahu SK, and Carlin BP (1995). Efficient parametrisations for normal linear mixed models. *Biometrika* 82, 479–488.

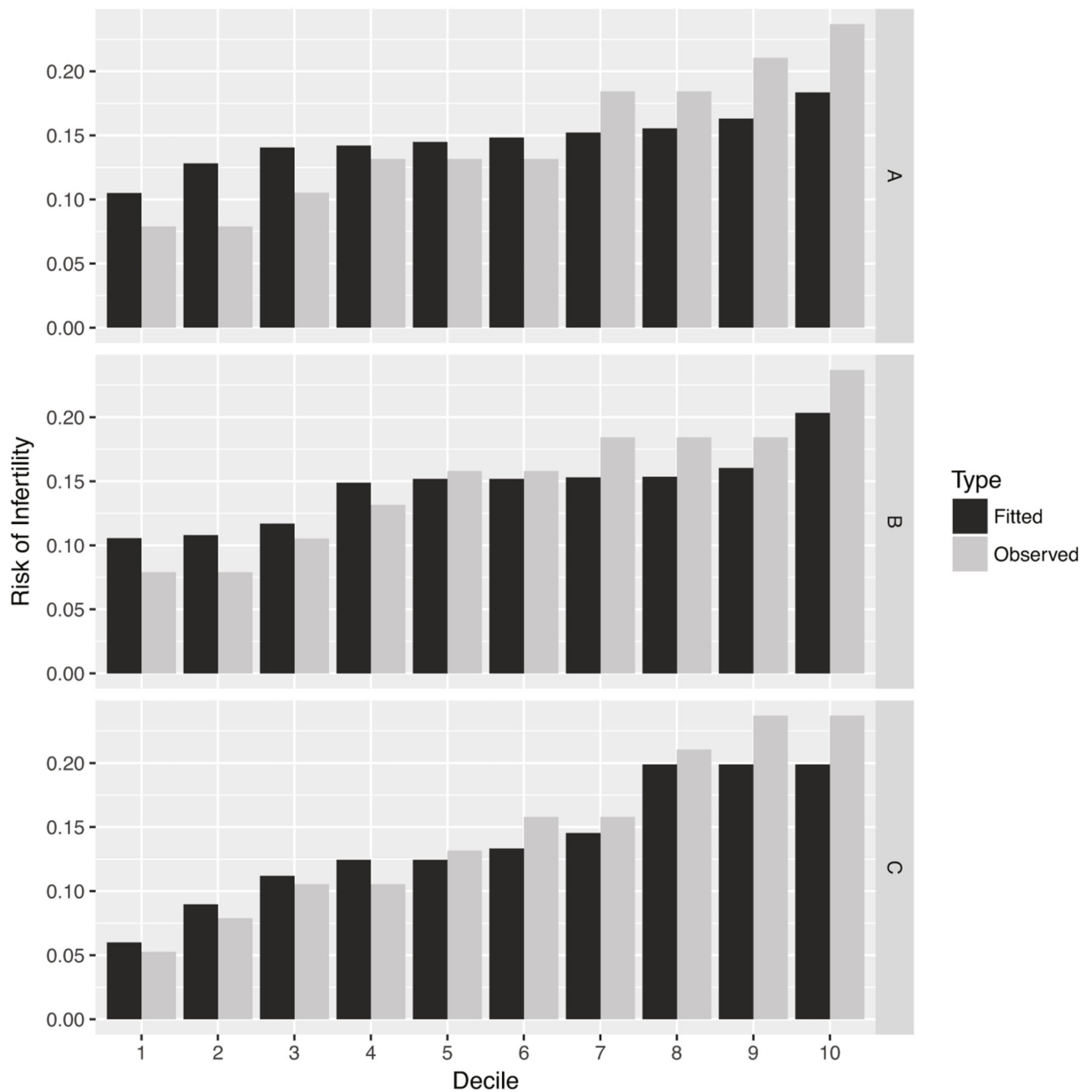
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, and Rubin DB (2014). *Bayesian Data Analysis*. Florida: Chapman and Hall.
- Haario H, Saksman E, and Tamminen J. (2005). Component-wise adaptation for high dimensional MCMC *Comput Stat* 20, 265–273.
- Lenk PJ and DeSarbo WS (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* 65, 93–119.
- Lin H, McCulloch CE, Turnbull BW, and Slate EH (2002). Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal prostate-specific antigen readings and prostate cancer. *Am Stat Assoc* 97, 53–65.
- Liu Y, Liu L, and Zhou J. (2015). Joint latent class model of survival and longitudinal data: An application to CPCRA Study. *Comput Stat Data Anal* 91, 40–50.
- Liu L, Strawderman RL, Cowen ME, and Shih YCT (2010). A flexible two-part random effects model for correlated medical costs. *J Health Econ* 29, 110–123. [PubMed: 20015560]
- Liu L, Strawderman RL, Johnson BA, and O'Quigley JM (2012) Analyzing repeated measures semi-continuous data, with application to an alcohol dependence study. *Stat Methods Med Res* 0, 1–20.
- Manning WG, Morris CM, Newhouse JP, et al. (1981). A Two-Part Model of the Demand for Medical Care: Preliminary Results from the Health Insurance Study. In *Health, Economics, and Health Economics*, van der Gaag J and Perlman M. (eds), North Holland: Amsterdam.
- Neelon B, O'Malley AJ, and Normand S-LT (2011) A Bayesian two-part latent class model for longitudinal medical expenditure data: Assessing the impact of mental health and substance abuse parity. *Biometrics* 67, 280–289. [PubMed: 20528856]
- Practice Committee of the American Society for Reproductive Medicine. (2013). Definitions of infertility and recurrent pregnancy loss: A committee opinion *Fertil Steril* 99, 63.
- Spiegelhalter DJ, Best NG, Carline BP, and Linde A. (2002). Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc Ser B* 64, 583–639.
- Su L, Tom BDM, and Farewell VT (2009). Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics* 10, 374–389. [PubMed: 19136448]
- Watanabe S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res* 11, 3571–3594.
- Zhang B, Chen Z, and Albert PS (2012). Latent class models for joint analysis of disease prevalence and high-dimensional semicontinuous biomarker data. *Biostatistics* 13, 74–88. [PubMed: 21908867]

**FIGURE 1.**

Histograms of PCB congener 153 in females and males in the LIFE Study. Original scales are in (a) and (b) and log scales in (c) and (d)

**FIGURE 2.**

Barplots of the estimated class membership probabilities in the adjusted joint models with different numbers of risk classes in the LIFE Study. The black color indicates the probability of belonging to the highest risk class for each model.

**FIGURE 3.**

Fitted and observed risk of infertility by deciles in the LIFE Study. “A” refers to the model with 5 latent risk classes for both partners where ordinal latent variables are used, “B” refers to the model with 3 latent risk classes for both partners where ordinal latent variables are used, and “C” refers to the model with 3 latent risk classes for both partners where indicator latent variables are used.

TABLE 1

Descriptive statistics of covariates on the original scale in the LIFE Study.

| Covariate | Female | | | | | Male | | | | |
|------------------------|--------------|--------------------|------------------|----------------|----------------|-------|-------|----------------|----------------|----------------|
| | Mean | SD | Q ₁ | Q ₂ | Q ₃ | Mean | SD | Q ₁ | Q ₂ | Q ₃ |
| Age (years) | 29.8 | 4.0 | 27 | 29 | 33 | 31.7 | 4.7 | 28 | 31 | 35 |
| Serum lipids (ng/g) | 616.3 | 115.9 | 530.0 | 603.1 | 677.5 | 731.6 | 216.3 | 592.4 | 687.4 | 803.2 |
| Serum cotinine (ng/mL) | 14.0 | 59.61 | 0.008 | 0.02 | 0.05 | 48.0 | 131.4 | 0.014 | 0.036 | 0.32 |
| Covariate | Level | Female N(%) | Male N(%) | | | | | | | |
| BMI | <25 | 186(49.2) | 69(18.3) | | | | | | | |
| (kg/m ²) | 25–30 | 97(25.7) | 151(39.9) | | | | | | | |
| | 30 | 95(25.1) | 158(41.8) | | | | | | | |

TABLE 2
Estimated DICs in the adjusted joint model with different numbers of classes in the LIFE Study

| No. of Classes | Males | | | | | | | |
|----------------|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| Females | 2 | -154854.4 | -155840.2 | -156309.4 | -156533.8 | -156824.1 | -157213.2 | -157298.7 |
| | 3 | -155773.8 | -156761.7 | -157286.4 | -157472.6 | -157590.2 | -157702.8 | -157669.9 |
| | 4 | -156332.2 | -157338.5 | -157850.1 | -158121.7 | -158328.4 | -158352.8 | -158112.3 |
| | 5 | -156550.0 | -157524.8 | -158048.8 | -158296.9 | -158437.0 | -158470.1 | -158520.0 |
| | 6 | -156788.5 | -157610.1 | -158239.9 | -158501.3 | -158625.3 | -158691.6 | -158479.1 |
| | 7 | -157009.1 | -157795.5 | -158281.6 | -158609.4 | -158669.5 | -158793.6 | -158511.8 |
| | 8 | -157113.2 | -157803.0 | -158092.2 | -158577.8 | -158501.1 | -158552.5 | -158647.3 |

TABLE 3

Adjusted joint model in the LIFE Study: Posterior means and 95% credible intervals for parameters.

| Parameter | 2-class model | 3-class model | 4-class model | 5-class model |
|------------------|---------------------|---------------------|---------------------|---------------------|
| β_0 | -1.99(-2.39, -1.64) | -2.12(-2.59, -1.71) | -2.30(-2.92, -1.79) | -2.32(-2.96, -1.77) |
| β_1^F | 0.59(0.14, 1.20) | 0.46(0.11, 0.97) | 0.41(0.11, 0.83) | 0.34(0.09, 0.70) |
| β_1^M | 0.48(0.11, 1.04) | 0.46(0.11, 0.91) | 0.44(0.12, 0.88) | 0.47(0.13, 0.91) |
| β_2 | -1.14(-2.32, -0.10) | -0.53(-1.06, -0.06) | -0.31(-0.62, -0.05) | -0.27(-0.52, -0.06) |
| α_h^F | -5.91(-6.29, -5.53) | -5.97(-6.35, -5.58) | -6.20(-6.59, -5.81) | -6.27(-6.66, -5.89) |
| α_1^F | 0.70(0.60, 0.80) | 0.57(0.48, 0.66) | 0.46(0.38, 0.54) | 0.41(0.34, 0.48) |
| α_0^M | -5.54(-5.93, -5.14) | -5.65(-6.04, -5.25) | -5.79(-6.18, -5.40) | -5.83(-6.22, -5.44) |
| α_1^M | 0.63(0.53, 0.73) | 0.56(0.48, 0.64) | 0.45(0.37, 0.53) | 0.43(0.35, 0.51) |
| η_0^{F*} | 1.80(1.72, 1.87) | 1.78(1.71, 1.86) | 1.81(1.73, 1.88) | 1.82(1.74, 1.90) |
| η_1^F | 1.84(1.76, 1.92) | 1.80(1.72, 1.88) | 1.82(1.74, 1.90) | 1.83(1.75, 1.92) |
| η_h^{M*} | 2.22(2.13, 2.30) | 2.20(2.12, 2.28) | 2.26(2.18, 2.35) | 2.27(2.18, 2.36) |
| η_1^M | 1.81(1.73, 1.89) | 1.77(1.69, 1.85) | 1.83(1.75, 1.91) | 1.84(1.76, 1.92) |
| τ_F^2 | 0.28(0.27, 0.29) | 0.25(0.24, 0.25) | 0.23(0.23, 0.24) | 0.22(0.22, 0.23) |
| τ_M^2 | 0.29(0.28, 0.30) | 0.26(0.25, 0.26) | 0.24(0.24, 0.25) | 0.24(0.23, 0.24) |
| σ_{0F}^2 | 1.35(0.84, 2.15) | 1.37(0.85, 2.19) | 1.38(0.86, 2.20) | 1.33(0.83, 2.13) |
| σ_{1F}^2 | 0.09(0.05, 0.14) | 0.07(0.04, 0.11) | 0.05(0.03, 0.08) | 0.05(0.03, 0.07) |
| σ_{1M}^2 | 1.43(0.89, 2.28) | 1.43(0.89, 2.28) | 1.39(0.86, 2.21) | 1.39(0.87, 2.21) |
| ρ_{01}^{FF} | 0.26(-0.08, 0.56) | 0.18(-0.15, 0.49) | 0.09(-0.24, 0.41) | 0.11(-0.23, 0.42) |
| ρ_{00}^{FM} | 0.97(0.94, 0.98) | 0.97(0.94, 0.98) | 0.97(0.94, 0.98) | 0.97(0.94, 0.98) |
| ρ_{01}^{FM} | 0.19(-0.15, 0.49) | 0.10(-0.23, 0.41) | 0.11(-0.21, 0.42) | 0.10(-0.24, 0.42) |

| Parameter | 2-class model | 3-class model | 4-class model | 5-class model |
|------------------|---------------------|--------------------|--------------------|--------------------|
| F_{10}^M | 0.32(−0.01,0.60) | 0.23(−0.10,0.53) | 0.14(−0.18,0.45) | 0.16(−0.17,0.47) |
| ρ_{11}^{FM} | 0.55(0.29,0.75) | 0.45(0.14,0.68) | 0.39(0.08,0.64) | 0.35(0.04,0.61) |
| ρ_{01}^{MM} | 0.25(−0.08,0.54) | 0.15(−0.18,0.46) | 0.16(−0.17,0.47) | 0.15(−0.18,0.45) |
| λ_1^F | 0.15(0.14,0.16) | 0.13(0.11,0.15) | 0.11(0.09,0.13) | 0.07(0.05,0.11) |
| λ_2^F | 0.11(0.08,0.14) | 0.06(0.002,0.10) | 0.16(0.12,0.19) | 0.16(0.12,0.20) |
| λ_3^F | −0.0002(−0.03,0.03) | 0.03(−0.08,0.02) | 0.001(−0.09,0.07) | −0.05(−0.08,−0.01) |
| λ_4^F | 0.03(0.01,0.04) | 0.07(0.06,0.08) | 0.03(0.01,0.04) | 0.07(0.05,0.09) |
| λ_5^F | −0.01(−0.02,−0.002) | −0.03(−0.06,−0.01) | −0.01(−0.03,0.01) | 0.004(−0.03,0.02) |
| λ_1^M | 0.16(0.14,0.18) | 0.10(0.09,0.12) | 0.09(0.07,0.11) | 0.10(0.09,0.12) |
| λ_1^M | −0.12(−0.16,−0.08) | −0.23(−0.29,−0.19) | −0.18(−0.21,−0.14) | −0.13(−0.17,−0.09) |
| λ_3^M | −0.09(−0.13,−0.05) | −0.17(−0.22,−0.13) | −0.07(−0.11,−0.04) | −0.04(−0.08,−0.01) |
| λ_4^M | 0.17(0.15,0.18) | 0.20(0.19,0.22) | 0.12(0.10,0.13) | 0.11(0.10,0.13) |
| λ_5^M | −0.03(−0.05,−0.02) | −0.13(−0.15,−0.11) | −0.10(−0.11,−0.08) | −0.11(−0.12,−0.09) |
| π_0^F | 0.67(0.62,0.72) | 0.55(0.49,0.60) | 0.30(0.24,0.37) | 0.19(0.14,0.23) |
| π_1^F | − | 0.36(0.31,0.42) | 0.40(0.34,0.47) | 0.39(0.34,0.45) |
| π_2^F | − | − | 0.23(0.19,0.27) | 0.27(0.23,0.32) |
| π_3^F | − | − | − | − |
| π_0^M | 0.69(0.64,0.74) | 0.44(0.39,0.50) | 0.31(0.26,0.37) | 0.29(0.24,0.34) |
| π_1^M | − | 0.43(0.38,0.49) | 0.42(0.37,0.48) | 0.42(0.36,0.47) |
| π_2^M | − | − | 0.20(0.16,0.24) | 0.22(0.18,0.26) |
| π_3^M | − | − | − | 0.05(0.03,0.08) |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

| Parameter | 2-class model | 3-class model | 4-class model | 5-class model |
|-----------|---------------|---------------|---------------|---------------|
| DIC | -154854.4 | -156761.7 | -157850.1 | -158296.9 |