



NIH PUBLIC ACCESS

Author Manuscript

Epidemiology. Author manuscript; available in PMC 2012 September 20.

Published in final edited form as:

Epidemiology. 2010 July ; 21(Suppl 4): S71–S76. doi:10.1097/EDE.0b013e3181cf0058.

Nonparametric Bayes Shrinkage for Assessing Exposures to Mixtures Subject to Limits of Detection

Amy H. Herring

Department of Biostatistics and Carolina Population Center, The University of North Carolina at Chapel Hill

Abstract

Assessing potential associations between exposures to complex mixtures and health outcomes may be complicated by a lack of knowledge of causal components of the mixture, highly correlated mixture components, potential synergistic effects of mixture components, and difficulties in measurement. We extend recently proposed nonparametric Bayes shrinkage priors for model selection to investigations of complex mixtures by developing a formal hierarchical modeling framework to allow different degrees of shrinkage for main effects and interactions and to handle truncation of exposures at a limit of detection. The methods are used to shed light on data from a study of endometriosis and exposure to environmental polychlorinated biphenyl congeners.

Introduction

Assessing the relationship between exposures x_1, \dots, x_p and a health outcome y may be complicated by multiple factors, including the following: a lack of understanding of the biologic pathways through which particular mixture elements, or interactions among them, may affect human health; near collinearity among a potentially large number p of mixture components; and difficulties in measuring mixture elements. In the absence of existing knowledge about the active components of a mixture, investigators often turn to model fitting and selection techniques to determine whether mixture components are associated with outcomes. Standard techniques, however, are not designed to identify the causal component in a mixture. For example, suppose one is interested in a binary health outcome y . One standard approach involves fitting the fully adjusted multiple logistic regression model

$$\text{logit}(Pr(y_i=1|x_{i1}, \dots, x_{ip})) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}, \quad (1)$$

where β_0 is the baseline log-odds of $y_i = 1$, and β_j is the increase in the log-odds attributable to a unit increase in the j th exposure.

When moderate-to-high correlations exist among the x_j , as in the case for many mixtures, the estimates of β and the accompanying estimated standard errors may become inflated because of instabilities caused by near collinearity. Standard model selection techniques, such as backward or forward selection, may lead to different models and different conclusions about active components in the mixture. After model selection, results from a final model are reported with accompanying confidence intervals or P -values. Such an approach ignores the uncertainty that arises from the selection and results in confidence intervals that are too narrow. To acknowledge model uncertainty, Bayesian model averaging^{1,2} can be used. Wang and colleagues³ have shown improved performance for Bayesian model averaging over stepwise selection methods in logistic regression. Neither

Bayesian nor frequentist⁴ model averaging has been widely applied in observational epidemiology, due in part, to a lack of implementation of these methods in mainstream software packages.

Another common approach is to consider a model containing an exposure summary, such as

$$\text{logit}(Pr(y_i=1|x_{i1}, \dots, x_{ip}))=\beta_0+\beta_1 \sum_{j=1}^p w_j x_{ij}, \quad (2)$$

in which the weights w_j may be set to $\frac{1}{p}$ in which a simple average over all components is used, set to 1 so that an exposure sum is used, or estimated in advance by using a method such as principal components. However, these summary approaches generally do not identify particular active elements in a mixture, and the results may be difficult to interpret.

When randomized trials or animal bioassay studies are appropriate, study design principles may be used to shed light on active elements of a mixture. In particular, fixed-ratio ray designs have been used to assess the impact of subsets of mixture components on disease risk.^{5,6} These designs have many applications in animal bioassay studies, but their applicability to human studies, in which exposures to chemical mixtures may not be ethical, is limited.

Numerous authors have suggested using hierarchical models as a means of overcoming model instability when a fully adjusted model is used.⁷⁻¹² Consider a study that examines the association between endometriosis and exposure to polychlorinated biphenyls (PCBs). Although numerous PCB congeners have been identified, we have relatively little information from humans to suggest whether all congeners, a single congener, or certain combinations of congeners confer excess risk of endometriosis. For example, consider the model

$$\text{logit}(Pr(y_i=1|x_{i1}, \dots, x_{ip}))=\beta_0+\sum_{j=1}^p \beta_j x_{ij}+\sum_{j=1}^p \sum_{k=1}^{j-1} \gamma_{jk} x_{ij} x_{ik}, \quad (3)$$

where γ_{jk} , $j = 1, \dots, p$, $k = 1, \dots, j-1$ are parameters for 2-way interactions between the p congeners.

When using maximum likelihood (ML) estimation, we encounter two problems in this setting. For highly correlated predictors, the maximum likelihood estimates (MLEs) are often unstable with large variance. The second problem is dimensionality: as p grows, the total number of parameters to be estimated rapidly becomes large. Investigators have proposed a variety of approaches that rely on penalized ML.^{13,14} Penalized ML relies on maximizing the log-likelihood plus a penalty that favors coefficients close to zero. Commonly used approaches rely on L_1 or L_2 penalties. The L_1 approach penalizes the sum of the absolute values of the regression coefficients (and interaction coefficients, if appropriate), and the L_2 approach penalizes the squared coefficients. The lasso procedure,¹³ which has been widely used for accommodating high-dimensional and correlated predictors, relies on an L_1 penalty. Methods that rely on an L_2 penalty are typically referred to as ridge regression approaches. Genkin et al.¹⁴ developed a lasso procedure for high-dimensional logistic regression and compared it with ridge logistic regression.

Lasso and ridge logistic regression shrink the coefficient estimates toward zero, which leads to stabilized estimation and to lower mean squared error than unpenalized ML. As the

amount of information in the data increases, the likelihood will become dominant, and the penalty will play a decreasing role. These methods are mainly designed for cases in which the data do not contain sufficient information to estimate the parameters precisely. When the number of parameters to estimate, including coefficients for main effects and interactions, is moderate to large, shrinkage can lead to substantial improvement in performance even when the sample size is not small. For example, ML estimates of 100 parameters that are based on a few thousand subjects are often poorly behaved relative to shrinkage estimates. An appealing property of lasso relative to ridge regression is simultaneous shrinkage and variable selection because many of the estimated coefficients can be exactly zero in the lasso approach. If the estimated coefficient is zero for an exposure, this exposure is effectively excluded from the model.

Penalized ML approaches have a Bayesian interpretation, with the penalty arising through a prior distribution. Under a double-exponential prior distribution for the coefficients, the mode of the posterior distribution is exactly equal to the lasso estimate. The ridge regression estimator arises through independent normal priors on the coefficients. For Bayesian robustness, it is common practice to use a prior with heavy tails, such as a t -distribution with low degrees of freedom or a Cauchy distribution. The nice feature of heavy-tailed priors is robustness to misspecification of the prior mean. We may choose the prior mean to be zero because it is likely that most of the exposure coefficients in a high-dimensional case are close to zero, but there may be a few outliers corresponding to important exposures having coefficients that are not close to zero. The Cauchy distribution does a good job of characterizing such a case because it shrinks coefficients for exposures with biologically insignificant odds ratios (ORs) strongly towards zero while limiting shrinkage of important exposures. Such priors are advocated by Gelman et al.¹⁵ Cauchy priors seem preferable to double-exponential priors, which have light tails and may lead to overshrinkage of coefficients for important exposures. However, Cauchy priors do not automatically lead to variable selection. In the machine learning literature, it is common to rely on t priors with zero degrees of freedom,¹⁶ with such priors having heavy tails and a variable selection property.

Variable selection resulting from the double-exponential or t prior with zero degrees of freedom occurs because the estimates can be exactly zero. However, no measure of uncertainty exists in this variable selection. Considering model (3), an alternative Bayesian approach chooses a variable selection mixture prior for the coefficients.^{17,18} The prior distribution for β_j consists of a mixture of 2 components, with 1 component (the spike) concentrated at zero and the other (the slab) with a relatively high variance. The spike represents the coefficients for predictors having ORs close to 1, whereas the slab provides a prior distribution on the coefficients for the important exposures. This approach accomplishes simultaneous variable selection and shrinkage, as in the lasso. Unlike in the lasso, however, we also fully account for uncertainty in selection. From Markov chain Monte Carlo (MCMC) samples, we can estimate the proportion of times that a given coefficient (say β_j) falls in the slab. This statistic provides an estimate of the posterior probability that the j th exposure has a main effect in the model and accounts for uncertainty in the other exposures and interactions included in the model. From the samples of β_p , we can also estimate a 95% credible interval (CI), which is a Bayesian version of the confidence interval. Similar analyses can be conducted for the interaction coefficients.

A potential concern is sensitivity to the choice of distribution for the slab. A standard choice is a normal distribution with moderate-to-large variance. However, this may be unrealistic in studies in which exposure ORs that are far from 1 are seldom encountered, so that a prior distribution that places small probabilities on ORs larger than 10 or smaller than 0.1 is usually preferred. In addition, performance in variable selection and model averaging can be

adversely impacted if the variance of the normal is chosen to be very large. In this case, a tendency exists to select small models that have potentially important exposures removed. The centering of the slab on zero (corresponding to an OR=1) may be undesirable; if effects are not null, estimates should not be close to zero. A multimodal prior that concentrates mass at locations away from zero is more realistic. Mixture distributions that contain components representing strong adverse effects, mild positive effects, and other reasonable associations are a natural solution. However, the number and location of the mixture components are not known in advance. Nonparametric Bayes shrinkage priors¹⁰ have recently been shown to have desirable properties in identifying important elements of a large group of predictors, with estimates closer to the truth on average compared with those from other hierarchical models.¹¹

Measuring PCB exposures is a challenging issue because many may be below the limit of detection (LOD). Statistical issues associated with analysis of data subject to an LOD are well known.^{19,20} Censored observations are a type of informative missing data because investigators typically know that with high probability the true observed values lie below the LOD. In practice point imputation of values such as zero or the LOD divided by the square root of 2 is often used, but these single imputation approaches lead to underreporting of variability in point estimates because the uncertainty in the exact values of exposure measurements that fell below the LOD is not taken into account. Using MCMC algorithms,²¹ repeated imputation of missing values, based on sampling from conditional distributions incorporating information about the LOD, may be used to account for uncertainty in exposure values below the LOD.

Methods

Study Characteristics

The study population is described in detail elsewhere.²² Briefly, 84 women aged 18–40 years who were undergoing laparoscopic investigation were enrolled in the study and asked to provide a blood sample. Upon laparoscopic examination, 32 women had confirmed endometriosis and 52 did not. Gas chromatography with electron capture was used to quantify levels of 62 PCB congeners in serum samples from 79 women.²³ The affiliated university and participating hospitals gave institutional review board consent for this study, and consent was obtained from all the women. We analyze 7 PCBs (94, 114, 142, 149, 153, 167, and 206) for which more than 50% of participants had detectable levels of the congeners. The majority of the PCBs are below the LOD for all but a small percentage of the participants, so there is essentially no information in the data about the relationships between such PCBs and disease risk. We focused on a subset in which substantial numbers of participants had data above the LOD. The 7 PCBs have low-to-moderately high correlation. The highest correlation is in the pairs (114, 167), (114, 153), (167, 153), and (206, 153); all the pairs have correlation coefficients between 0.6 and 0.7.

Statistical Methods

We square-root transformed and then normalized PCBs before analyzing the data. Transforming exposures subject to LOD is not problematic as long as the transformation is monotone and the detection threshold is also transformed. By normalizing the exposures, we simplified the selection of a default shrinkage prior that did not need to take into account the measurement scale of the data. Such normalization is standard practice in applying shrinkage procedures. The logistic regression model (3) was used for the probability of endometriosis (y_i) conditional on main effects and 2-way interactions between PCBs, with a total of 7 main effects and 21 interaction terms of interest.

The goals of the Bayesian approach were 3-fold: 1) accommodate model uncertainty by using model averaging, 2) use shrinkage priors to reduce mean square estimation error, and 3) accommodate truncation of exposure data at the LOD. A shrinkage prior¹⁰ for the regression coefficients β and γ was used to address goals 1) and 2). We used a Dirichlet process prior distribution with a model selection component. This specification allowed us to place substantial prior probability on the value zero, which represents no association of a PCB with endometriosis. This prior distribution has nice properties in reducing the MSE. We defined separate prior distributions for main effects and interactions that placed a greater prior probability on zero values for interaction terms than for zero values on main effects. These prior distributions were specified as follows:

$$\begin{aligned}
 \beta_j &\sim D_1, j=1, \dots, 7 \\
 D_1 &\sim DP(\lambda_1 D_{01}) \\
 D_{01} &= \pi_1 \delta_0 + (1-\pi_1) N(\mu_1, \varphi_1^2) \\
 \gamma_{jk} &\sim D_2, j=1, \dots, 7, k=1, \dots, j-1 \\
 D_2 &\sim DP(\lambda_2 D_{02}) \\
 D_{02} &= \pi_2 \delta_0 + (1-\pi_2) N(\mu_2, \varphi_2^2),
 \end{aligned} \tag{4}$$

where δ_0 represents a point mass at zero. To adjust for multiple testing in Bayesian procedures, one can choose a hierarchical prior on the probability of exclusion for each predictor.²⁴ In our model, we have a component for main effects and another for interactions. We let $\pi_1 \sim \text{beta}(1, 1)$ be the prior for the main effect exclusion probabilities. This prior has expectation $E(\pi_1) = 0.5$, which corresponds to an equal chance of inclusion or exclusion of an exposure in the main effect component. However, as we add unimportant predictors, this prior will be updated to be centered closer and closer to 1, which induces a multiplicity adjustment. If only a single exposure is important, it will be less likely to be judged as significant if many unimportant exposures are present.

For the interaction inclusion probability, we chose $\pi_2 \sim \text{beta}(9, 1)$, which has expectation $E(\pi_2) = 0.9$. By centering the prior closer to 1, we favor a sparser structure in the interaction component of the model, with a larger adjustment for multiplicity. This specification reflects prior knowledge that relatively few interactions may exist compared with main effects. The amount of information in a beta prior can be measured in terms of the prior sample size, which corresponds to the sum of the 2 beta parameters. Hence, our beta(9, 1) prior corresponds to direct data from 10 interaction coefficients, with 9 of these being zero and the remaining nonzero. This beta prior will be updated with the allocation to the zero or nonzero components for the 21 interaction coefficients in the model. As 21 is substantially larger than 10, it may naively seem that the data play more of a role than the prior and that an insufficient penalty may exist through the prior to favor a sparse structure. However, it is important to keep in mind that the allocation to the zero and non-zero components is unknown, so that the effective sample size of the data in terms of Bayesian learning about π_2 is much less than 21. Finally, to complete the prior specification, the intercept β_0 was given a $N(0, 4)$ prior distribution. This seems to provide a reasonable choice of shrinkage prior given that the exposure data have been normalized before analysis.

To address LOD, we modeled square-root transformed PCB values by using normal distributions, with the PCB values falling below the LOD modeled as truncated normal (TN). The TN distribution is simply a normal distribution restricted to a subset of the real line. The TN distribution provides an adequate fit to the data falling above the LOD. Let \mathbf{c} be a $p \times 1$ vector of (possibly transformed) LOD for the variables of interest. Then a general specification of the joint distribution, as a product of marginal and conditional distributions, is given as follows:

$$(x_{i1}, \dots, x_{ip}) \sim N(\mathbf{x}'_{i[6]} \alpha_7, \tau_7^{-1}, c_7) \times N(\mathbf{x}'_{i[5]} \alpha_6, \tau_6^{-1}, c_6) \times \dots \times N(\mathbf{x}'_{i[1]} \alpha_2, \tau_2^{-1}, c_2) \times N(\alpha_1, \tau_1^{-1}(\mathfrak{z})_1),$$

in which $\mathbf{x}'_{i[k]} = (x_{i1}, \dots, x_{ik})$. After normalizing the \mathbf{x} values, the prior distributions for the regression coefficients are given standard multivariate normal distributions $\alpha_k \sim N(\mathbf{0}, \mathbf{I})$, and the precisions are given gamma priors $\tau_k \sim \text{Gamma}(1, 1)$, $k = 1, \dots, p$. For the PCB data, only 2 PCB's had values below the LOD, so that we model only these two components of \mathbf{x} .

We obtained estimates by using Gibbs sampling^{25,26} in the software package Matlab (Mathworks, version 2007b, Natick, MA) using a data augmentation Gibbs sampler. The Matlab code is available from the authors upon request. The Gibbs sampler produces draws from the posterior distribution of the coefficients for the main effects and interactions. The code is efficient; it required less than 1 minute to collect 5000 samples for the PCB example. To assess the performance, we ran a simulation study chosen to mimic the PCB data. We assumed that all main effects and interactions were zero, with the exception of $\beta_3 = 1$, $\beta_6 = 1$ and $\gamma_{46} = 1$. We simulated 100 data sets that had the same data structure as the PCB data and analyzed each using i) the proposed Bayes nonparametric shrinkage approach, ii) ML logistic regression, and iii) ridge logistic regression. For approach i), we collected 5000 Gibbs iterations, with the first 1000 discarded as a burn-in to allow for convergence. This number was sufficient for convergence based on examination of trace plots. For approaches ii) and iii), we plugged in the LOD divided by the square root of 2 for the exposures under the LOD. We used MSE averaged across the different parameters and 100 simulated data sets to assess performance and obtained the following MSEs: i) 0.116; ii) 1.210; and iii) 0.208. Hence, ML logistic regression had extremely poor performance, with ridge logistic regression providing a dramatic improvement. The proposed approach further improved upon ridge logistic regression by a substantial amount.

Results

In this study, 48% of values of PCB 94 were below the LOD and 44% of values of PCB 153 were below the LOD. All other PCBs considered had 100% of values above the LOD. After a burn-in of 3000 samples, we based inference on 10,000 samples from the data augmentation Gibbs sampler. Visual examination of trace plots provided no evidence of lack of convergence. The posterior inclusion probabilities, which provided a posterior estimate of the probability that a main effect or interaction effect is associated with endometriosis, were for the most part quite small (Table 1). However, we found substantial evidence (posterior inclusion probability 0.99) that higher levels of PCB 114 exposure were associated with increased endometriosis (estimated OR = 3.7; 95% CI = (1.6–8.2) for a 1-standard deviation increase in exposure). The results were robust to moderate changes to our prior specification based on sensitivity analyses in which we changed the hyperparameter values to several different reasonable alternative values and reran the analysis.

Although a frequentist multiple logistic regression model was fit to the data for comparison purposes, we found that near collinearity caused instabilities in estimates, and the model did not converge. Hence, as a reasonable alternative, we fit logistic regression models containing only a main effect for a single exposure at a time and added pairwise interactions for those exposures with significant main effects; we used a threshold of 0.05 on the P -values. In this analysis, we observed that only PCB 114 was significant (P -value = 0.0003). The same conclusion was also obtained using a logistic ridge regression approach. Although these simpler approaches produced the same overall conclusion as the proposed nonparametric Bayes approach, this does not provide evidence that the proposed approach

does not have substantial practical advantages. The PCB application provides a single example, and the advantages have been compellingly illustrated in the simulation study.

Discussion

It is commonplace to collect data with a moderate-to-large number of predictors. Identifying the important predictors from the many candidates is a challenging problem. The number of possible subsets of p predictors is 2^p , which grows extremely rapidly with p . Hence, even in relatively small problems having $p = 20$ or 30 , there are enormous numbers of possible subsets of important predictors. Traditional methods of variable selection, such as stepwise selection, fail in this setting by not adequately exploring the huge space of possible models or accounting for uncertainty in the model selection process. Bayesian variable selection solves this problem by allowing one to search the model space, while simultaneously estimating marginal inclusion probabilities for each predictor. These probabilities provide a weight of evidence that a predictor should be included and adjust for uncertainty in other predictors in the model. The Bayesian approach is computationally intensive, but this computational burden is necessary to explore the different models and accurately accommodate uncertainty. Although the literature on Bayesian variable selection is abundant, almost all the approaches rely on hierarchical models in which the distribution on the coefficients to be included is a unimodal distribution centered on zero.

In this article, we advocate the use of flexible mixture distributions, which can more realistically characterize variability in the exposure ORs in epidemiologic studies that involve mixtures of correlated exposures. Implementation can proceed with an MCMC algorithm, which involves iteratively sampling from standard distributions. This approach can easily be implemented in R or Matlab.

Correlated predictors present a major challenge in statistical analysis. In addition to the issues with multicollinearity, almost all variable selection methods have a tendency to include only 1 of a set of correlated predictors in the model. The reason for this tendency is that if one has multiple correlated exposures and wants to select a model that does a good job in parsimoniously predicting the response, the best approach may be simply to take 1 of the exposures. Standard stepwise selection and modern variable selection procedures based on lasso and relevance vector machine methods are all subject to this problem. The issue is that the statistical procedure is designed for prediction instead of inferences. Some recent literature has focused on cleverly defining penalization regions to select simultaneously groups of correlated predictors that are important.²⁷ When predictors are very highly correlated, Bayesian stochastic search variable selection algorithms, such as the approach proposed in this article, can also face challenges. For example, if 2 important exposures are very highly correlated, the marginal inclusion probabilities may be close to 0.5 instead of close to 1. Hence, it tends to be the case that very highly correlated predictors are more difficult to detect, although one can examine probabilities of including a pair of predictors in an attempt to bypass this problem.

References

1. Draper D. Assessment and propagation of model uncertainty. *J R Stat Soc Series B*. 1995; 57:45–97.
2. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Stat Sci*. 1999; 14:382–417.
3. Wang DL, Zhang WY, Bakhai A. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. *Stat Med*. 2004; 23:3451–3467. [PubMed: 15505893]
4. Hjort NL, Claeskens G. Frequentist model average estimators. *J Am Stat Assoc*. 2003; 98:879–899. (with discussion).

5. Casey M, Gennings C, Carter WH, Moser VC, Simmons JE. Detecting interaction(s) and assessing the impact of component subsets in a chemical mixture using fixed ratio mixture ray designs. *J Agri Biol Environ Stat.* 2004; 9:339–361.
6. Casey M, Gennings C, Carter WH, Moser V, Simmons JE. D-optimal designs for studying combinations of chemicals using multiple fixed-ratio ray experiments. *Environmetrics.* 2005; 16:129–149.
7. Greenland S. A semi-Bayes approach to the analysis of correlated multiple associations, with an application to an occupational cancer mortality study. *Stat Med.* 1992; 11:219–230. [PubMed: 1579760]
8. Greenland S. Methods for epidemiologic analyses of multiple exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. *Stat Med.* 1993; 12:717–736. [PubMed: 8516590]
9. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. *Environ Health Persp.* 1994; 102(Suppl 8):33–39.
10. Dunson DB, Herring AH, Engel SM. Bayesian selection and clustering of polymorphisms in functionally related genes. *J Am Stat Assoc.* 2008; 103:534–546.
11. MacLehose RF, Dunson DB, Herring AH, Hoppin JA. Bayesian methods for highly correlated exposure data. *Epidemiology.* 2007; 18:199–207. [PubMed: 17272963]
12. Thomas DC. Viewpoint: using gene-environment interactions to dissect the effects of complex mixtures. *J Expos Sci Environ Epidemiol.* 2007; 17:S71–S74.
13. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc, Series B.* 1996; 58:267–288.
14. Genkin A, Lewis DD, Madigan D. Large-scale Bayesian logistic regression for text categorization. *Technometrics.* 2007; 49:291–304.
15. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat.* 2008; 2:1360–1383.
16. Tipping ME. Sparse Bayesian learning and the relevance vector machine. *J Machine Learning Res.* 2001; 1:211–244.
17. Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *J Am Stat Assoc.* 1988; 83:1023–1032.
18. George EI, McCulloch RE. Approaches to Bayesian variable selection. *Stat Sinica.* 1997; 7:339–373.
19. Krishnamoorthy K, Mallick A, Matthew T. Model-based imputation approach for data analysis in the presence of non-detects. *Ann Occup Hyg.* 2009; 53:249–263. [PubMed: 19181626]
20. Taylor DJ, Kupper LL, Rappaport SM, Lyles RH. A mixture model for occupational exposure mean testing with a limit of detection. *Biometrics.* 2001; 57:681–688. [PubMed: 11550915]
21. Robert, CP.; Casella, G. *Monte Carlo Statistical Methods.* New York: Springer-Verlag; 1999.
22. Buck Louis GM, Weiner JM, Whitcomb BW, et al. Environmental PCB exposure and risk of endometriosis. *Human Reprod.* 2005; 20:279–285.
23. Whitcomb BW, Schisterman EF, Buck GM, Weiner JM, Greizerstein H, Kostyniak PJ. Relative concentrations of organochlorine pesticides and polychlorinated biphenyls in adipose tissue and serum of women of reproductive age. *Environ Toxicol Pharmacol.* 2005; 19:203–213. [PubMed: 21783478]
24. Scott JG, Berger JO. An exploration of aspects of Bayesian multiple testing. *J Stat Plann Inference.* 2006; 136:2144–2162.
25. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Analysis Machine Intell.* 1984; 6:721–741.
26. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc.* 1990; 85:398–409.
27. Bondell HD, Reich BJ. Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR. *Biometrics.* 2008; 64:115–123. [PubMed: 17608783]

Table 1

Estimated log ORs, 95% CIs, and Posterior Inclusion Probabilities for Associations Between PCB Exposures and Endometriosis

PCB	Estimated log OR (95% CI)	Posterior inclusion probability
94	0.0 (-0.2, 0.4)	0.19
114	1.3 (0.5, 2.1)	0.99
142	0.1 (0.0, 0.9)	0.28
149	0.1 (-0.1, 0.6)	0.23
153	0.0 (-0.7, 0.4)	0.22
167	0.0 (-0.4, 0.4)	0.21
206	0.0 (-0.3, 0.5)	0.20
94 × 114	0.0 (0.0, 0.3)	0.05
94 × 142	0.0 (-0.2, 0.0)	0.05
94 × 149	0.0 (-0.3, 0.0)	0.05
94 × 153	0.0 (-0.7, 0.0)	0.07
94 × 167	0.0 (0.0, 0.4)	0.06
94 × 206	0.0 (-0.3, 0.0)	0.06
114 × 142	0.0 (0.0, 0.0)	0.05
114 × 149	0.0 (0.0, 0.0)	0.03
114 × 153	0.0 (0.0, 0.0)	0.03
114 × 167	0.0 (0.0, 0.1)	0.04
114 × 206	0.0 (0.0, 0.0)	0.05
142 × 149	0.0 (0.0, 0.5)	0.06
142 × 153	0.0 (-0.7, 0.0)	0.07
142 × 167	0.0 (-0.1, 0.0)	0.04
142 × 206	-1.3 (-2.5, 0.0)	0.91
149 × 153	0.0 (0.0, 0.2)	0.05
149 × 167	0.0 (-0.1, 0.0)	0.04
149 × 206	0.0 (0.0, 0.0)	0.04
153 × 167	0.0 (-0.2, 0.0)	0.04
153 × 206	0.0 (-0.3, 0.0)	0.05
167 × 206	-0.1 (-0.9, 0.0)	0.10

ORs, odds ratios; CI, credible interval