Bayesian distributed lag interaction models to identify perinatal windows of vulnerability in children’s health

ANDER WILSON∗
Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA
ander.wilson@colostate.edu

YUEH-HSIU MATHILDA CHIU
Department of Environmental Medicine and Public Health & Kravis Children’s Hospital,
Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

HSIAO-HSIEN LEON HSU
Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai,
New York, NY 10029, USA

ROBERT O. WRIGHT
Department of Environmental Medicine and Public Health & Kravis Children’s Hospital,
Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

ROSALIND J. WRIGHT
Department of Environmental Medicine and Public Health & Kravis Children’s Hospital,
Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

BRENT A. COULL
Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

SUMMARY
Epidemiological research supports an association between maternal exposure to air pollution during pregnancy and adverse children’s health outcomes. Advances in exposure assessment and statistics allow for estimation of both critical windows of vulnerability and exposure effect heterogeneity. Simultaneous estimation of windows of vulnerability and effect heterogeneity can be accomplished by fitting a distributed lag model (DLM) stratified by subgroup. However, this can provide an incomplete picture of how effects vary across subgroups because it does not allow for subgroups to have the same window but different within-window effects or to have different windows but the same within-window effect. Because the timing of some developmental processes are common across subpopulations of infants while for others the timing differs across subgroups, both scenarios are important to consider when evaluating health risks of prenatal exposures. We propose a new approach that partitions the DLM into a constrained functional predictor that estimates windows of vulnerability and a scalar effect representing the within-window effect directly.

∗To whom correspondence should be addressed.

© The Author 2017. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
The proposed method allows for heterogeneity in only the window, only the within-window effect, or both. In a simulation study we show that a model assuming a shared component across groups results in lower bias and mean squared error for the estimated windows and effects when that component is in fact constant across groups. We apply the proposed method to estimate windows of vulnerability in the association between prenatal exposures to fine particulate matter and each of birth weight and asthma incidence, and estimate how these associations vary by sex and maternal obesity status in a Boston-area prospective pre-birth cohort study.

**Keywords**: Birth weight; Child asthma; Distributed lag models; Exposure effect heterogeneity; Fine particulate matter; Functional data analysis.

### 1. Introduction

A growing body of research supports an association between maternal exposure to air pollution during pregnancy and a variety of birth and children’s health outcomes. Epidemiological studies have found that maternal exposure to ambient air pollution is associated with decreased birth weight as well as increased risk of preterm birth and respiratory disorders including asthma (Šrám and others, 2005; Kelly and Fussell, 2011; Shah and Balkhair, 2011; Stieb and others, 2012; Jedrychowski and others, 2013; Savitz and others, 2014; Chiu and others, 2014). The National Institute of Environmental Health Sciences (NIEHS) has identified both the estimation of windows of vulnerability and the identification of susceptible subpopulations as critical research directions in environmental health research (NIEHS, 2012). A critical methodological gap is the lack of available statistical methods to simultaneously identify windows of vulnerability and susceptible populations.

Windows of vulnerability are time periods during which exposure to a toxin has an increased association with current or future health status (Barr and others, 2000; West, 2002). Prenatal systems development is a multi-event process progressing sequentially from early gestation (Kajekar, 2007). The identification of windows of vulnerability, which in turn corresponds to sensitive stages of development, can inform our understanding of the underlying pathways through which an environmental exposure operates. Presumably, the window is defined by developmental specific events (i.e. gene expression changes, growth/cell density, vascularization etc.) that are transient and environmental exposure that is concurrent to these events.

The distributed lag model (DLM) framework has a long history in air pollution research and was originally developed for time-series analysis where an outcome observed on a given day is jointly regressed on exposures over a previous time period (Schwartz, 2000; Zanobetti and others, 2000). Several recent studies have applied DLMs to estimate windows vulnerability during which air pollution has an elevated association with preterm birth (Warren and others, 2012; Chang and others, 2015), decreased birth weight (Warren and others, 2013), childhood asthma (Hsu and others, 2015), and disrupted neurodevelopment (Chiu and others, 2016).

A critical consideration in the estimation of windows of vulnerability is that the prenatal developmental process is not homogeneous across all subgroups of individuals. For example, females display earlier fetal breathing than males (Becklake and Kauffmann, 1999). Because developmental timing varies, it is natural to hypothesize that the association between prenatal exposure and health outcomes in a given subgroup may not only vary in the effect size but in the timing of the window of vulnerability. Therefore, when interest focuses on windows of vulnerability, there are at least four potential patterns of effect heterogeneity: (i) both the effect size and the timing of the window vary by subgroup; (ii) only the effect size varies by subgroup; (iii) only the timing of the window varies by subgroup; or (iv) both the window and the effect size are the same for all subgroup. Existing methods only accommodate patterns (1) and (4), but not (2) and (3).
A DLM can estimate a common window and effect size (pattern 4) or, when stratified by group, estimate lagged effects assuming that both the effect size and the exposure window vary by subgroup (pattern 1), (e.g. Hsu and others, 2015; Chiu and others, 2016). The stratified DLM approach does not share information across groups relating to either the timing of the window or the effect size within the window. In the context of functional regression, Wei and others (2014) proposed a functional interaction model for gene-environment interactions with a time-varying environmental exposure. This approach shares information across groups (genotype) by assuming the same windows for each genotype but the effect is scaled by the number of major alleles (0, 1, or 2). Hence, this approach can estimate effects satisfying pattern 2 under the additional constraint that the effects are proportional to the number of major alleles. However, none of these approaches allow for the estimation across all four patterns of effect heterogeneity.

In this article, we propose a new method that provides greater flexibility in characterizing effect heterogeneity when identifying windows of vulnerability is of interest. The proposed Bayesian distributed lag interaction model (BDLIM) partitions the distributed lag function from a standard DLM into a time component that identifies windows of vulnerability and a scale component that quantifies the magnitude of the effect within the window. The approach allows both the window and the scale components to either vary or stay constant across subgroups. As such, the BDLIM framework can estimate a model with any of the four effect heterogeneity patterns. To our knowledge, models that assume the window of vulnerability is the same across subgroups but the effect within the window varies across groups, and vice versa, have not previously been considered in the literature. The proposed approach allows the user to directly answer the question of whether effect heterogeneity manifests itself via changes in the window of vulnerability, the magnitude of the effect, or both, which in turn more directly answers the question of whether an environmental exposure affects the developmental process of an infant similarly across subgroups.

In the BDLIM framework, the time-varying weight function is treated as a functional predictor that is scaled by the scalar effect size. This partitioning requires that identifiability constraints be placed on the parameters of the weight function. Under certain assumptions about effect heterogeneity the model can be reparameterized to relax the identifiability constraints on the parameter space and reduces to a mixed effects model. In other cases, including the general linear model setting for discrete responses, we use a slice sampler to efficiently estimate the model from the constrained space. We make software available for BDLIM in the R package regimes (REGression In Multivariate Exposure Settings).

We use BDLIM to estimate the association between fine particulate matter (PM2.5) measured weekly over pregnancy and two outcomes—birth weight for gestational age (BWGA) z-score and asthma incidence—in a prospective Boston-area pregnancy cohort. Following Lakshmanan and others (2015), we evaluate whether the association between PM2.5 and BWGA z-score varies by both sex and maternal obesity status. Following Hsu and others (2015), we evaluate how effects on asthma incidence vary by sex.

2. THE ACCESS DATA

We analyze data from the Asthma Coalition on Community, Environment, and Social Stress (ACCESS) project (Wright and others, 2008). ACCESS is a prospective, longitudinal study designed to investigate the effects of stress and other environmental factors, including air pollution, on asthma risk in an urban US setting. The ACCESS cohort includes data on 997 mother–child pairs that were recruited between August 2002 and January 2007. The women were at least 18 years of age, spoke English or Spanish, and received prenatal care at one of two Boston, MA area hospitals or affiliated community health centers. To date, the ACCESS cohort has been used to study the relationship of air pollution exposures, maternal stress, and other risk factors with outcomes including asthma, wheeze, and birth weight (Chiu and others, 2012; Hsu and others, 2015; Lakshmanan and others, 2015). Like previous ACCESS
studies, we limit the analysis to full-term (≥37 weeks), live births with complete exposure, outcome, and covariate data.

For each child we consider BWGA z-scores and maternal-reported and clinician-diagnosed asthma as outcomes. The data contain maternal and child covariate information including: maternal age at enrollment; race/ethnicity (black, hispanic, and white); maternal education (two categories less than high school and high school diploma or more); self-reported smoking during pregnancy; indicator of maternal prepregnancy obesity; infant sex; maternal atopy (ever self-reported doctor-diagnosed asthma, eczema, or hay fever); season of birth; a previously described maternal stress index (Chiu and others, 2012); and a previously described neighborhood disadvantage index (Chiu and others, 2012). Maternal exposures of PM$_{2.5}$ were estimated with a hybrid land use regression model that incorporates satellite-derived aerosol optical depth measures (Kloog and others, 2011). Each mother was assigned an average PM$_{2.5}$ exposure value for each week of pregnancy based on the predicted value at her address of residence. We limit our analysis to exposure during the first 37 weeks of pregnancy.

3. Bayesian distributed lag interaction models

3.1. Approach

Interest focuses on estimation of the association between time-varying exposure $X_i(t)$, $t \in T$, and scalar outcome $Y_i$ while controlling for a vector of baseline covariates $Z_i$. We denote individuals by $i = 1, \ldots, n$.

We parameterize the time-varying effect of exposure as $\beta w(t)$, where $w(t)$ is a continuous weight function that captures temporal variation in the association between $X(t)$ and $Y$ and $\beta$ is the scalar effect size. The weight function $w(t)$ identifies windows of vulnerability in which the exposure effect is elevated relative to other time periods. To allow for heterogeneity among subgroups (e.g. infant sex) indexed by $j$, we allow either or both of these quantities to vary across levels of $j$, denoted $\beta_j$ and $w_j(t)$. Similar to a stratified DLM, this parameterization allows for group-level modification of both the window and effect size. In addition, this parameterization accommodates scenarios not yet considered whereby only the location of the window or only the magnitude of the effect, but not both, vary by group.

3.2. BDLIM for a single population

In the BDLIM, regression model for a single population with no effect heterogeneity we assume $E(Y_i) = \mu_i$ and

$$g(\mu_i) = \alpha + \beta \int_{t \in T} X_i(t)w(t)\,dt + Z_i^T \gamma,$$

(3.1)

where $g(\cdot)$ is a monotone link function, $\alpha$ is the intercept, and $\gamma$ is a vector of unknown regression coefficients for the covariates $Z$. The model in (3.1) is similar to a functional linear model in which the total effect of exposure $X(t)$ on outcome $Y$ at time $t$ is $\beta w(t)$. We will refer to this model as BDLIM-n, where the “n” indicates no heterogeneity between subgroups.

For identifiability, we constrain the weight function such that $\int_{t \in T} w(t)^2\,dt = 1$ and $\int_{t \in T} w(t)\,dt \geq 0$. The weight function is allowed to be both positive and negative to account for exposures that are a toxicant during some time periods but a nutrient during others. The constraint $\int_{t \in T} w(t)\,dt \geq 0$ assures that $\beta$ reflects the direction of the cumulative effect, $\beta \int_{t \in T} w(t)\,dt$. The constraint $\int_{t \in T} w(t)^2\,dt = 1$ ensures that the magnitude of $\beta$ (i.e. $|\beta|$) is identifiable.

To gain some insight into the BDLIM approach consider the case where $w(t) = 1$. In this case, $\int_{t \in T} X_i(t)w(t)\,dt = \bar{X}_i$ and (3.1) becomes a linear model with scalar exposure covariate equal to the mean exposure over the full pregnancy. However, once the weight function $w(t)$ varies with $t$, the weighted
Bayesian distributed lag interaction models

3.3. BDLIM model for effect heterogeneity

A key advantage of the BDLIM framework is the ability to estimate the three hypothesized patterns of heterogeneity where either $\beta$, $w(t)$, or both $\beta$ and $w(t)$ vary by group. When either $\beta$ or $w(t)$ is constant across groups, BDLIM yields a more parsimonious model that results in more powerful tests of an interaction.

Consider analysis of data for groups $j = 1, \ldots, J$. When both the effect size and the window of vulnerability are group-specific, the BDLIM model is

$$g(\mu_i) = \alpha_{ji} + \beta_{ji} \int_{t \in T} X_i(t)w_{ji}(t)dt + Z_i^T \gamma,$$

(3.2)

where the subscript $ji$ denotes group $j$ to which individual $i$ belongs. We refer to this model as BDLIM-bw where “bw” indicates that both $\beta$ and $w(t)$ vary across groups.

The BDLIM framework extends to the new scenarios where only $\beta$ or $w(t)$ varies by group. If it is hypothesized that groups share a common window, e.g. first trimester, but the groups are differentially susceptible within that window, the model is

$$g(\mu_i) = \alpha_{ji} + \beta \int_{t \in T} X_i(t)w(t)dt + Z_i^T \gamma,$$

(3.3)

which we denote BDLIM-b. Alternatively, if the effect is the same across groups but the windows are different, perhaps shifted by a few weeks, then the model, denoted BDLIM-w, is

$$g(\mu_i) = \alpha_{ji} + \beta \int_{t \in T} X_i(t)w_{ji}(t)dt + Z_i^T \gamma.$$

(3.4)

Both BDLIM-b and BDLIM-w estimate patterns of effect heterogeneity not previously addressed.

3.4. Parameterization of the functional components

We assume a truncated basis function representation of both $X_i(t)$ and $w(t)$. Common choices for the basis expansions include splines, wavelets, Fourier series, and principal components (PCs). We use the first $K$ PCs of the covariance matrix of $X(t)$ as the basis to represent both $\theta_{ji}(t)$ and $w(t)$ for all groups. Here, $K$ is chosen to be the number of PCs that explain a prespecified proportion of the total variability in the exposures, for example 99% of the total variation. Hence, $\theta_{ji}(t) = \sum_{k=1}^K \xi_{ik} \psi_k(t)$ and $w(t) = \sum_{\ell=1}^K \theta_{\ell} \psi_\ell(t)$.

In practice, we observe the exposures measured over a discrete grid, in our case $t = 1, \ldots, T$, of pregnancy. Let $X$ be a $n \times T$ matrix with row $i$ being the observed exposures for individual $i$ measured at times $t = 1, \ldots, T$, $X_i = (X_{i1}, \ldots, X_{iT})$. We use the covariance matrix of $X$ to estimate the basis functions $\{\psi_k(t)\}_{k=1}^K$. It is reasonable to expect that $w(t)$ is moderately smooth. However, the raw PCs of the covariance matrix of $X$ are potentially rough. To obtain a smooth orthonormal basis, we use fast covariance estimation (FACE) proposed by Xiao and others (2016) to obtain the eigenfunctions of a smoothed covariance matrix, as implemented in the R packagerefund (Crainiceanu and others, 2014). There are several potential alternative approaches. In Appendix C of the supplementary materials available at Biostatistics online, we consider pre-smoothing the exposures and using the PCs of the covariance of...
Additionally, we complete the model by assigning prior distributions to the unknown regression parameters. The presented as proportional to a constrained multivariate standard normal, \( \theta \), parameter space as described in Section 3.6.

The orthonormal PC basis facilitates implementation of the constraints on \( w(\cdot) \). When \( w(t) = \sum_{k=1}^{K} \theta_k \psi_k(t) \), the constraint \( \int_{\mathbb{R}} w(t)^2 dt = 1 \) is satisfied if and only if \( \|\theta\| = 1 \), where \( \theta = (\theta_1, \ldots, \theta_T)^T \).

For the normal linear model, BDLIM-n and BDLIM-bw can be reparameterized to reduce the computational burden imposed by the constrained parameter space. For BDLIM-n, the unknown parameters \( \theta \) for the weight functions \( w(\cdot) \) are constrained to \( \|\theta\| = 1 \) and \( 1^T \Psi \theta \geq 0 \); however, \( \beta \) is unconstrained and \( \theta^* \equiv \beta \theta^* \) is also unconstrained in \( \mathbb{R}^K \). Importantly, \( \beta = \|\theta^*\| \times \text{sign}(1^T \Psi \theta^*) \), \( \theta = \theta^* \|\theta^*\|^{-1} \), and \( \hat{w} = \Psi \theta \) are each uniquely identified from \( \theta^* \). This generalizes to the BDLIM-bw with group specific \( \theta^j \).

### 3.5. Reparameterization to remove constraints in the linear model

For the normal linear model, BDLIM-n and BDLIM-bw can be reparameterized to reduce the computational burden imposed by the constrained parameter space. For BDLIM-n, the unknown parameters \( \theta \) for the weight functions \( w(\cdot) \) are constrained to \( \|\theta\| = 1 \) and \( 1^T \Psi \theta \geq 0 \); however, \( \beta \) is unconstrained and \( \theta^* \equiv \beta \theta^* \) is also unconstrained in \( \mathbb{R}^K \). Importantly, \( \beta = \|\theta^*\| \times \text{sign}(1^T \Psi \theta^*) \), \( \theta = \theta^* \|\theta^*\|^{-1} \), and \( \hat{w} = \Psi \theta \) are each uniquely identified from \( \theta^* \). This generalizes to the BDLIM-bw with group specific \( \theta^j \).

### 3.6. Prior specification and computation

We complete the model by assigning prior distributions to the unknown regression parameters. The prior for \( \theta \) is uniform over the surface of the unit \( K \)-hemiball for \( \theta \). The prior likelihood can be represented as proportional to a constrained multivariate standard normal, \( \pi(\theta) \propto \exp(-\theta^T \theta/2)1\{\|\theta\| = 1\}1\{1^T \Psi \theta \geq 0\} \), where \( 1\{\cdot\} \) is an indicator function.

We assign normal priors to \( \beta \) and \( \gamma \). When reparameterized, \( \beta \) becomes a scale parameter for \( \theta^* \).

Specifically, we assume that \( \beta \sim N(0, \tau^2) \) and let \( \kappa = \beta^2 \tau^{-2} \). Then \( \theta^* \sim N(0, \kappa^{-1/2}I) \), where \( \tau \) is fixed and \( \kappa \sim \chi^2_1 \). The resulting model has closed form full conditionals: \( \theta^* \) can be sampled from a multivariate normal and \( \kappa \) from a generalized inverse-Gaussian distribution with density function \( f(\kappa; \lambda, \chi, \psi) \propto \kappa^{(n-2)/2} \exp[-(\chi/\kappa + \psi\kappa)/2] \), where \( \lambda = -(K - 1)/2 \), \( \psi = 1 \), and \( \chi = \tau^{-2}(\theta^*^T \theta^*)^{-1} \).

Finally, we assume a flat prior on the intercept \( \alpha \) and, for the linear model with residuals \( \epsilon_i \sim N(0, \sigma^2) \), a gamma prior on the precision parameter \( \sigma^2 \).

We use MCMC to simulate the posterior of the unknown parameters. For the BDLIM-n and BDLIM-bw in the linear model, all unknown parameters have simple conjugate forms and can be sampled via Gibbs sampler. For BDLIM-b, BDLIM-w, and models with a non-linear link function, we propose a slice sampling approach (Neal, 2003) based on the elliptical slice sampler proposed by Murray and others (2010) (see Appendix B of the supplementary materials available at Biostatistics online for details).
We quantify the evidence in the data supporting each of the four potential patterns of effect heterogeneity. We tested the performance of BDLIM with two simulation studies. Simulation A compares the BDLIM to pre-smooth the exposures instead of smoothing the covariance matrix of X with FACE, in Appendix C.

3.7. Summarizing the posterior of w(t)
Summarizing the posterior distribution of w(t) deserves special consideration in light of the constraint \( \int_{t_1}^{t_2} w(t)^2 dt = 1 \). It is typical to summarize the posterior distribution with the posterior mean. However, the posterior mean of \( \theta \) almost surely does not satisfy \( \| \theta \| = 1 \) and \( \int_{t_1}^{t_2} w(t)^2 dt = 1 \). To obtain a point estimate for w(t) such that \( \int_{t_1}^{t_2} w(t)^2 dt = 1 \) and \( \int_{t_1}^{t_2} w(t)dt \geq 0 \), we take the posterior mean of \( \theta \) in the topology of a K-hemiball parameter space. We do this by taking the Bayes estimate with respect to the loss function \( L(\theta, \tilde{\theta}) = \frac{1}{(\| \theta - \tilde{\theta} \|^2)} \). The resulting estimate \( \hat{\theta} \) is the posterior mean projected onto the K-hemiball. That is, \( \hat{\theta} = \theta \| \theta \|^{-1} \), where \( \theta \) is the posterior mean. Since each draw from the posterior satisfies \( 1^T \Psi \theta \geq 0 \), it follows that \( 1^T \Psi \hat{\theta} \geq 0 \). We identify windows of vulnerability as time periods where the pointwise 95% posterior intervals of \( \hat{w}(t) \) do not contain 0.

A drawback to this point estimate is that in the absence of an effect (\( \beta = 0 \)) or when \( \theta \) is not well identified by the data, the posterior of \( \theta \) will reflect the prior. In this case, the projected estimator can be erratic; however, the posterior interval will still reflect the uncertainty in \( \hat{w}(t) \).

The effect \( \beta \) is interpretable even when a window is not identified. Because \( \int_{t_1}^{t_2} w(t)dt \geq 0 \), the effect \( \beta \) and the cumulative effect \( \beta \int_{t_1}^{t_2} w(t)dt \) share the same significance level, i.e. \( \Pr(\beta > 0 | D) = \Pr(\beta \int_{t_1}^{t_2} w(t)dt > 0 | D) \). Hence, the \( \alpha \)-level posterior interval for the cumulative effect will not contain 0 if and only if the \( \alpha \)-level posterior interval for \( \beta \) does not contain 0 and, regardless of whether a window is identified, we can conclude that there is an overall effect.

3.8. Comparing models and identifying the pattern of heterogeneity
We quantify the evidence in the data supporting each of the four potential patterns of effect heterogeneity with the mean log posterior predictive density (MLPPD). Specifically, for model k (where \( k = 1, 2, 3, 4 \) indicates the four BDLIM variants) \( \text{MLPPD}_k = S^{-1} \sum_{s=1}^{S} \log \Pr(Y|s^{(s)}, \zeta_k) \), where \( \zeta_k \) is the vector of all parameters for model k and \( s = 1, \ldots, S \) enumerates the draws from the simulated posterior distribution. We compute \( \text{MLPPD}_k \) for each of the four effect heterogeneity models, then normalize as

\[
\hat{P}_k = \frac{\exp \left( \text{MLPPD}_k \right)}{\sum_{i=1}^{4} \exp \left( \text{MLPPD}_i \right)}, \quad (3.6)
\]

In the simulation study presented in Section 4.3, we compare the performance of computed using the normalized MLPPD to identify the correct model to that of the deviance information criterion (DIC, Spiegelhalter and others, 2002). Note that DIC is \(-2\text{MLPPD} + p_D\), where \( p_D \) is an additional penalty for model size. We show that normalized MLPPD is less likely to identify a misspecified pattern of effect heterogeneity.

4. Simulation
4.1. Simulation overview
We tested the performance of BDLIM with two simulation studies. Simulation A compares the BDLIM-\( \phi \) to a DLM when interest focuses on estimation of the effect of a time-varying exposure in a single group. Results suggest that the two methods perform similarly. We have relegated most of the details for this simulation to Appendix C of the supplementary materials available at Biostatistics online. We also compared tuning choices for BDLIM, including using different numbers of PCs and using natural splines to pre-smooth the exposures instead of smoothing the covariance matrix of X with FACE, in Appendix C.
of the supplementary materials available at Biostatistics online. Simulation B highlights the advantage of BDLIM for subgroup analyses and effect heterogeneity estimation.

For both simulations, we used the observed exposures and covariates from the birth weight analysis of the ACCESS data. Using observed weekly air pollution levels during pregnancies ensures that the exposure data have realistic temporal trends and autocorrelations. Hence, the data contain \( n = 506 \) individuals with exposures measured at 37 evenly spaced time points, 10 binary covariates, 3 continuous covariates, and an intercept. For the second simulation, we divided the data into two groups, 239 girls \( (j = 0) \) and 267 boys \( (j = 1) \).

For each scenario, we simulated 1000 datasets and analyzed them with BDLIM and DLM. For BDLIM, we used 15 knots to estimate the covariance matrix. We used \( N(0, 10^2) \) priors on \( \beta \) and \( \gamma \) and used the first \( K \) PCs that explain 99% of the variation in exposure. We assumed Gaussian errors and put a gamma \((0.01, 0.01)\) prior on \( \sigma^{-2} \). For DLM we used natural cubic splines with flat priors on the regression coefficients. We fit the DLM with degrees of freedom ranging from 3 to 10 and have presented results only from the best performing model.

4.2. Simulation A: Comparison to DLM with no effect heterogeneity

This simulation compares BDLIM-n with DLM when there is no effect heterogeneity. Both models are correctly specified but use different parameterizations and basis functions.

We used three scenarios each with data simulated from a different weight function:

\[
w^1(t) = \sqrt{\frac{t(t-1)^4}{B(5, 5)}} \tag{4.1}
\]
\[
w^2(t) = \frac{\sin(t\pi - \pi/4)}{\sqrt{T^{-1} \sum_{k=1}^{T} \sin(t_k\pi - \pi/4)^2}} \tag{4.2}
\]
\[
w^3(t) = 1, \tag{4.3}
\]

where \( B(\cdot, \cdot) \) is a beta function and \( t \) is scaled to the unit interval. The superscripts identify the weight functions and correspond to the three scenarios in simulation A. The intercept and regression coefficients for the covariates are simulated as standard normals and we generated independent normal residuals with zero mean and standard deviation of six. Figure 1 shows the weight functions and the first 100 estimated weight functions using BDLIM-n.

![Fig. 1. Estimated weight functions \( \hat{w}(t) \) for simulation A. The grey lines show the estimated weight functions from BDLIM-n for the first 100 datasets. The thick black dashed line is the true weight functions.](image-url)
Both BDLIM-n and DLM can estimate the total time-varying effect $\beta w(t)$, while only BDLIM-n individually identifies $\beta$ and $w(t)$. For this reason, we focused the comparison on estimation of $\beta w(t)$ and the cumulative effect $\sum_{t=1}^{T-1} \beta w(t)$. Table 1 supplementary material available at Biostatistics online shows that the models performed similarly. The models had similar model fit as measured by DIC, had similar RMSE and coverage near 95% for the estimate of $\beta w(t)$. For the cumulative effect, $\sum_{t=1}^{T} \beta w(t)$, we found that the bias and RMSE were similar for the methods and that both had posterior interval coverage near 95%. Hence, when there is no subgroup-specific analysis there is no information lost by using BDLIM instead of DLM.

4.3. Simulation B: Performance with effect heterogeneity

The second simulation scenario compares BDLIM using the four parameterizations for exposure effect heterogeneity. The five simulation scenarios are as follows:

1. B.1: $w_1(t) = w^1(t)$, $\beta_1 = 0.1$, $w_2(t) = w^1(t)$, $\beta_2 = 0.1$, no heterogeneity.
2. B.2: $w_1(t) = w^1(t)$, $\beta_1 = 0.1$, $w_2(t) = w^1(t)$, $\beta_2 = -0.2$, heterogeneity in $\beta$ only.
3. B.3: $w_1(t) = w^1(t)$, $\beta_1 = 0.1$, $w_2(t) = w^1(t)$, $\beta_2 = 0.0$, one group with no effect.
4. B.4: $w_1(t) = w^1(t)$, $\beta_1 = 0.1$, $w_2(t) = w^2(t)$, $\beta_2 = 0.2$, heterogeneity in $w(t)$ and $\beta$.
5. B.5: $w_1(t) = w^1(t)$, $\beta_1 = 0.1$, $w_2(t) = w^2(t)$, $\beta_2 = 0.1$, heterogeneity in $w(t)$ only.

For each simulation scenario, Table 1 shows the model fits from each model using normalized MLPPD, as described in Section 3.8, and DIC. For each of the five scenarios, there is more than one correctly specified model because BDLIM-bw is correctly specified under any of the four forms of effect heterogeneity. Use of MLPPD to identify the best fitting model (indicated with a * in Table 1) selects one of the correctly specified models with very high probability (at least 95% for all scenarios) and almost never selects a misspecified model. In contrast, DIC selects the simplest, correctly specified model (shown in bold in Table 1) at a higher rate but also identifies a misspecified model at a higher rate for Scenarios B.4 and B.5. Therefore, we conclude that use of MLPPD is a more conservative choice in that it selects a misspecified model at a lower rate at the cost of less power to rule out the most general BDLIM-bw model.

Table 2 summarizes inference for $\beta$ and $w(t)$ for the different BDLIM models. Overall, using the model that matches the pattern of heterogeneity (indicated by bold in Table 2) in the data provides the best inference. This is most notable for Scenario B.4 where both $\beta_i$ and $w_j(t)$ vary by group and BDLIM-bw provides accurate inference while the other approaches have greater bias, larger RMSE, and lower coverage for both $\beta$ and $w(t)$. Similarly, for Scenarios B.2 and B.3, BDLIM-b results in improved estimation of $w(t)$ by sharing information across groups. To a lesser extent, BDLIM-w yields $\beta$ estimates with lower RMSE in Scenario B.5.

In summary, BDLIM-n is nearly identical to a standard DLM when there is no heterogeneity across groups. When there is heterogeneity across groups, using BLDIM we can identify a correctly specified model with high probability. When that model is a reduced model, BDLIM results in improved inference of the weight function and effect size.

5. Analysis of prenatal air pollution exposure

5.1. Impact of sex and maternal obesity on air pollution effects on birth weight

We used BDLIM to estimate the association between PM$_{2.5}$ and BWGA z-score. Following the analysis of Lakshmanan and others (2015), we estimated this association by child sex and maternal obesity status. Of the 506 children with complete data including BWGA z-score, there were 155 females with non-obese mothers, 182 males with non-obese mothers, 84 females with obese mothers, and 85 males with obese...
### Table 1. Comparison of model fit with four BDLIM parameterizations for simulation B

<table>
<thead>
<tr>
<th>Scenario</th>
<th>B.1</th>
<th>B.2</th>
<th>B.3</th>
<th>B.4</th>
<th>B.5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean log posterior predictive distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b</td>
<td>0.30*</td>
<td>0.61*</td>
<td>0.64*</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>BDLIM-bw</td>
<td>0.21*</td>
<td>0.39*</td>
<td>0.32*</td>
<td>0.97*</td>
<td>0.42*</td>
</tr>
<tr>
<td>BDLIM-n</td>
<td>0.24*</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>BDLIM-w</td>
<td>0.24*</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.50*</td>
</tr>
<tr>
<td><strong>Proportion model selected using mean log posterior predictive distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b</td>
<td>0.26*</td>
<td>0.63*</td>
<td>0.68*</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>BDLIM-bw</td>
<td>0.13*</td>
<td>0.37*</td>
<td>0.28*</td>
<td>0.98*</td>
<td>0.28*</td>
</tr>
<tr>
<td>BDLIM-n</td>
<td>0.30*</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>BDLIM-w</td>
<td>0.31*</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.65*</td>
</tr>
<tr>
<td><strong>Mean DIC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b</td>
<td>3270.32*</td>
<td>3270.85*</td>
<td>3269.77*</td>
<td>3287.60</td>
<td>3279.42</td>
</tr>
<tr>
<td>BDLIM-bw</td>
<td>3274.57*</td>
<td>3275.51*</td>
<td>3273.16*</td>
<td>3276.02*</td>
<td>3274.93*</td>
</tr>
<tr>
<td>BDLIM-n</td>
<td>3270.44*</td>
<td>3504.49</td>
<td>3303.76</td>
<td>3300.47</td>
<td>3281.10</td>
</tr>
<tr>
<td>BDLIM-w</td>
<td>3273.94*</td>
<td>3360.48</td>
<td>3283.67</td>
<td>3293.79</td>
<td>3274.31*</td>
</tr>
<tr>
<td><strong>Proportion model selected using DIC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b</td>
<td>0.31*</td>
<td>0.92*</td>
<td>0.85*</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>BDLIM-bw</td>
<td>0.02*</td>
<td>0.08*</td>
<td>0.14*</td>
<td>0.89*</td>
<td>0.12*</td>
</tr>
<tr>
<td>BDLIM-n</td>
<td>0.56*</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>BDLIM-w</td>
<td>0.11*</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>0.63*</td>
</tr>
</tbody>
</table>

The top panel shows the mean MLPPD across the 1000 simulated datasets. The second panel shows the proportion of times each parameterization ranked as the best fitting model based on MLPPD. The third panel shows the average DIC for each parameterization. The fourth panel shows the proportion of times each model was selected as the best fitting model based on DIC. Numbers in bold indicate that the model is the simplest model that is correctly specified while an asterisk (*) indicates that the model is correctly specified. Note that multiple models can be correctly specified and the BDLIM-bw is always correctly specified.

Mothers. Lakshmanan and others (2015) associated average PM$_{2.5}$ over the entire pregnancy with BWGA $z$-score. Here, we estimate the association using the four variations of BDLIM to identify windows of vulnerability and use MLPPD to select the best fitting model. We assumed a normal linear model and used the same priors as described for the simulation in Section 4.1 and set $K$ to explain 99% of the variance in $X(t)$.

The BDLIM-b model had the highest normalized MLPPD at 0.96. The other models were 0.02 for BDLIM-bw, 0.01 for BDLIM-w, and 0.01 for BDLIM-n. Hence, we present results from the BDLIM-b, which assumes a single weight functions $w(t)$ shared by all four groups but group specific $\beta_j$.

Figure 2a shows the estimated group-specific effects $\hat{\beta}_j$. The results are consistent with those reported by Lakshmanan and others (2015), with a negative association between PM$_{2.5}$ and BWGA $z$-score among boys with obese mothers but not in the other groups. In addition, the posterior probability of a pairwise difference between boys with obese mothers and the other three groups range from 0.93 to 0.97. The posterior difference for the pairwise comparisons between the three non-significant groups range from 0.50 to 0.84 suggesting little evidence of differences between those groups. Further, because the model is saturated we can perform an ANOVA decomposition on the posterior sample to investigate main effects of
Table 2. Simulation results for the inference on $\beta$ and $w(t)$ using BDLIM for simulation B

<table>
<thead>
<tr>
<th>Group</th>
<th>Inference for $\beta$</th>
<th>Inference for $w(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>Scenario B.1: No heterogeneity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b Female</td>
<td>0.000</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-b Male</td>
<td>-0.001</td>
<td>0.012</td>
</tr>
<tr>
<td>BDLIM-bw Female</td>
<td>0.000</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-bw Male</td>
<td>-0.001</td>
<td>0.012</td>
</tr>
<tr>
<td>BDLIM-n Female</td>
<td>0.005</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-n Male</td>
<td>-0.001</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-w Female</td>
<td>-0.001</td>
<td>0.010</td>
</tr>
<tr>
<td>BDLIM-w Male</td>
<td>-0.001</td>
<td>0.010</td>
</tr>
<tr>
<td>Scenario B.2: Heterogeneity in $\beta$ only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b Female</td>
<td>0.000</td>
<td>0.012</td>
</tr>
<tr>
<td>BDLIM-b Male</td>
<td>0.001</td>
<td>0.017</td>
</tr>
<tr>
<td>BDLIM-bw Female</td>
<td>0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>BDLIM-bw Male</td>
<td>0.007</td>
<td>0.013</td>
</tr>
<tr>
<td>BDLIM-n Female</td>
<td>-0.176</td>
<td>0.176</td>
</tr>
<tr>
<td>BDLIM-n Male</td>
<td>0.124</td>
<td>0.125</td>
</tr>
<tr>
<td>BDLIM-w Female</td>
<td>-0.285</td>
<td>0.285</td>
</tr>
<tr>
<td>BDLIM-w Male</td>
<td>0.015</td>
<td>0.018</td>
</tr>
<tr>
<td>Scenario B.3: One group (males) with no effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b Female</td>
<td>0.002</td>
<td>0.012</td>
</tr>
<tr>
<td>BDLIM-b Male</td>
<td>-0.002</td>
<td>0.012</td>
</tr>
<tr>
<td>BDLIM-bw Female</td>
<td>-0.010</td>
<td>0.014</td>
</tr>
<tr>
<td>BDLIM-bw Male</td>
<td>-0.006</td>
<td>0.020</td>
</tr>
<tr>
<td>BDLIM-n Female</td>
<td>-0.041</td>
<td>0.042</td>
</tr>
<tr>
<td>BDLIM-n Male</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td>BDLIM-w Female</td>
<td>-0.022</td>
<td>0.024</td>
</tr>
<tr>
<td>BDLIM-w Male</td>
<td>0.078</td>
<td>0.079</td>
</tr>
<tr>
<td>Scenario B.4: Heterogeneity in both $w(t)$ and $\beta$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b Female</td>
<td>0.006</td>
<td>0.018</td>
</tr>
<tr>
<td>BDLIM-b Male</td>
<td>-0.022</td>
<td>0.042</td>
</tr>
<tr>
<td>BDLIM-bw Female</td>
<td>0.003</td>
<td>0.012</td>
</tr>
<tr>
<td>BDLIM-bw Male</td>
<td>-0.010</td>
<td>0.021</td>
</tr>
<tr>
<td>BDLIM-n Female</td>
<td>0.048</td>
<td>0.049</td>
</tr>
<tr>
<td>BDLIM-n Male</td>
<td>-0.052</td>
<td>0.054</td>
</tr>
<tr>
<td>BDLIM-w Female</td>
<td>0.043</td>
<td>0.045</td>
</tr>
<tr>
<td>BDLIM-w Male</td>
<td>-0.057</td>
<td>0.058</td>
</tr>
<tr>
<td>Scenario B.5: Heterogeneity in $w(t)$ only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDLIM-b Female</td>
<td>0.003</td>
<td>0.013</td>
</tr>
<tr>
<td>BDLIM-b Male</td>
<td>-0.022</td>
<td>0.028</td>
</tr>
<tr>
<td>BDLIM-bw Female</td>
<td>0.000</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-bw Male</td>
<td>-0.007</td>
<td>0.018</td>
</tr>
<tr>
<td>BDLIM-n Female</td>
<td>-0.002</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-n Male</td>
<td>-0.002</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-w Female</td>
<td>-0.003</td>
<td>0.011</td>
</tr>
<tr>
<td>BDLIM-w Male</td>
<td>-0.003</td>
<td>0.011</td>
</tr>
</tbody>
</table>

For each scenario 1000 datasets were fit. The table shows the bias, RMSE, and 95% credible interval coverage for $\hat{\beta}$ and the RMSE and 95% credible interval coverage for $\hat{w}(t)$. The bold rows indicate the model that was most frequently selected as the MLPPD model for that simulation scenario.
Fig. 2. Estimated group specific effect sizes $\hat{\beta}_j$ (left panel) and weight function $\hat{w}(t)$ (right panel) using the BDLIM-b model for the BWGA $z$-score analysis.

sex and obesity as well as an interaction effect (see Appendix D of the supplementary materials available at Biostatistics online for details).

The estimated weight function $\hat{w}(t)$ (Figure 2b) shows a trend of increased vulnerability in the earlier part of pregnancy, approximately weeks 5–20. Although we do not identify a window with high probability, the estimated cumulative effect over the full pregnancy always has the same sign and significance level as $\hat{\beta}_j$. For male infants with obese mothers we estimate a cumulative effect of $-0.225$ with 95% credible interval $(-0.476, -0.001)$. Hence, the results are suggestive of a negative association between PM$_{2.5}$ exposures in early pregnancy and lower BWGA $z$-score among boys with obese mothers.

The estimated time-varying effects $\beta_j w(t)$ are presented in Appendix D of the supplementary materials available at Biostatistics online. Results from the BDLIM-bw model are presented there as well. Comparing the results using BDLIM-bw and BDLIM-b shows that the model with a shared weight function BDLIM-b is more suggestive of the location of the window of vulnerability and has, on average, 14% smaller posterior standard deviation for $\hat{\beta}_j$. Therefore, there is evidence that the magnitude of the effect varies across groups, but no evidence that the timing of the window varies across groups primarily because there is no effect in three of the four groups.

5.2. Sex-specific effects of prenatal air pollution on asthma incidence

We next used a logistic BDLIM to estimate sex-specific associations between PM$_{2.5}$ and childhood asthma incidence in the ACCESS cohort. Hsu and others (2015) analyzed these data by stratifying by sex and applying a standard DLM to data in each stratum. Here we assess whether the magnitude of the effect, the timing of the effect, or both vary by sex. This analysis included data from 544 births with complete data including asthma. Again, BDLIM-b was the best fitting model for the asthma analysis with normalized MLPPD of 0.43. The others were 0.21 for BDLIM-bw, 0.20 for BDLIM-n, and 0.17 for BDLIM-w. We describe the results from the BDLIM-b model.
Figure 3 shows the estimated weight function $\hat{w}(t)$ and the estimated group-specific effects $\hat{\beta}_j$ using BDLIM-b. Figure 3a shows a positive and statistically significant association between PM$_{2.5}$ exposure and asthma incidence in boys but not in girls. The weight function (Figure 3b) identifies a window of vulnerability in 13–21. Hence, PM$_{2.5}$ exposures during the 13–21 were positively associated with asthma among boys. This is comparable to the widows identified by Hsu and others (2015) which were 12–26 for boys using a sex-stratified analysis and 14–20 when testing sex differences. Figure 6 of the supplementary materials available at Biostatistics online shows the estimated sex-specific odds ratios for a 10$\mu$g/m$^3$ increase in PM$_{2.5}$, which peak around 1.25 for boys. The results using BDLIM-bw were very similar and are included in Appendix E of the supplementary materials available at Biostatistics online; however, BDLIM-b yielded posterior standard deviations of $\hat{\beta}_j$ that were 10% smaller than their BDLIM-bw counterparts.

6. DISCUSSION

In this paper we have proposed BDLIM as a new tool that can be used estimate effect heterogeneity in time-varying exposures. This addresses a critical methodological gap for simultaneous estimation of windows of vulnerability and identifying susceptible subpopulations. Specifically, BDLIM allows for estimation of effect heterogeneity when subgroups have a common window of vulnerability but different effects within the window (BDLIM-b) or when subgroups have the same effect size but in different windows (BDLIM-w). In these scenarios, the resulting estimates had reduced bias and RMSE, an advantage of pooling information across groups when the effects are not different. We demonstrated this advantage both in the simulation study and the data analysis.

The proposed approach partitions the time-varying effect into two components. The first is a constrained functional predictor that captures the temporal variation in the effect and identifies windows of vulnerability. The second component is a scalar effect size that quantifies the effect within the window. In some situations the constraint on the parameters of the weight function can be removed by reparameterizing
the model. In this case the scalar effect size becomes a scale parameter in the model with a generalized inverse-Gaussian full conditional, which allows for simulating the posterior with a Gibbs sampler. In other situations, including generalized linear models, we use a slice sampler to efficiently simulate the posterior of the contained parameters. The constraints of the weight function make summarizing the posterior of the weight function difficult because the posterior mean does not satisfy the constraints. We address this by using a point estimate that is the Bayes estimate with respect to a non-standard loss function specifically chosen to yield a posterior summary that satisfies the constraints. We then identify windows of vulnerability where the pointwise posterior interval does not contain zero.

We analyzed data from the ACCESS cohort on the association between prenatal exposure to PM$_{2.5}$ and both birth weight and asthma incidence. In both cases the BDLIM analyses suggested a common window of vulnerability but different effect sizes within the window for both outcomes. Hence, in both analyses the model providing the best fit to the data could not have been estimated using existing methods. Our results identified a window of 13–21 where PM$_{2.5}$ exposures were associated with increased asthma incidence in boys but not in girls. In the analysis of the birth weight data, there was strong evidence of a negative association between PM$_{2.5}$ in the earlier part of pregnancy and decreased BWGA $z$-score among boys born to obese mothers.

The proposed approach assumes that $\beta$ and/or $w(t)$ are the same for all groups or different for all groups. When there are more than two groups it may be of interest to understand if only certain pairs of groups share one or more component. In the BWGA $z$-score analysis we performed an \textit{a posteriori} ANOVA decomposition to investigate if the main effect of child sex or maternal obesity status is important. This could be done because the model for $\beta$ was saturated. Another simple approach would be to define groups based on preliminary analysis and rerun the model. In the BWGA $z$-score analysis, this would mean two groups: boys with obese mothers in one group and everyone else in another. A more sophisticated extension would be to extend the approach to directly model how covariates influence each component, such as modeling the weight functions as $w_j(t) = w_0(t) + w_{sex}(t) + w_{obese}(t)$.

Identifying susceptible populations and windows of vulnerability are critical areas of future research as highlighted in the NIEHS strategic plan (NIEHS, 2012). BDLIM provides an essential tool for simultaneous identification of susceptible populations and critical windows of vulnerability when estimating of the health effects of environmental exposures.

Supplementary material is available at http://biostatistics.oxfordjournals.org. Software for BDLIM is included in the R package regimes available at anderwilson.github.io/regimes/.

Acknowledgments

Conflict of Interest: None declared.

Funding

The ACCESS study has been supported by grants (R01 ES010932, R01 ES013744; U01 HL072494, and R01 HL080674 to W.R.J., PI). This work was supported by USEPA grant 834798 and NIH grants (ES020871, ES007142, CA134294, ES000002, P30 ES023515). This publication’s contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA.
REFERENCES


[Received March 30, 2016; revised November 01, 2016; accepted for publication January 13, 2017]