JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Johns Hopkins University, Dept. of Biostatistics Working Papers

7-12-2011

Reduced Bayesian Hierarchical Models: Estimating Health Effects of Simultaneous Exposure to Multiple Pollutants

Jennifer F. Bobb
Johns Hopkins Bloomberg School of Public Health, jfeder@jhsph.edu

Francesca Dominici
Harvard School of Public Health, fdominic@hsph.harvard.edu

Roger D. Peng
Johns Hopkins University, rpeng@jhsph.edu

Suggested Citation

Bobb, Jennifer F.; Dominici, Francesca; and Peng, Roger D., "Reduced Bayesian Hierarchical Models: Estimating Health Effects of Simultaneous Exposure to Multiple Pollutants" (July 2011). *Johns Hopkins University, Dept. of Biostatistics Working Papers.* Working Paper 230.
http://biostats.bepress.com/jhubiostat/paper230

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
Copyright © 2011 by the authors

1 Introduction

The US Environmental Protection Agency (EPA) estimated that thousands of premature deaths and hundred of thousands cases of illness may be avoided by reducing pollution (EPA, 2011). Most epidemiological studies of air pollution and health have estimated the health effects associated with ambient exposure to individual pollutants adjusting for exposure to other pollutants and confounders. However, National Research Council (NRC) has recently questioned whether the current approach of setting separate National Ambient Air Quality Standards (NAAQS) for each of the six criteria pollutants adequately protects population health, as this approach may greatly underestimate risk (NRC, 2004).To meet the challenges of the NRC recommendations, new statistical methods are needed to account for multiple exposures and their interactions.

Previous multisite time series studies of the health effects of air pollution have estimated risks associated with exposure to a single pollutant. Dominici et al. (2000) developed a two-stage Bayesian hierarchical model to combine information across locations on the association between daily changes of a given pollutant and daily changes in the health outcome, adjusted for other pollutants and confounders. This approach has been applied to several national US studies for estimating independent associations of various pollutants of epidemiologic interest with different health outcomes, including mortality and cardiovascular and respiratory emergency hospital admissions (Dominici et al., 2006; Bell et al., 2004; Peng et al., 2008, 2009).Two-level random-effect models have also been used to estimate health effects of exposure to individual pollutants and to identify factors that explain heterogeneity in the health risks across European cities (Katsouyanni et al., 2001).To address the potential for biased estimates due to measurement error of correlated exposures in multipollutant models, Zeka and Schwartz (2004) have applied methodology developed by Schwartz and Coull (2003) to estimate independent effects of individual pollutants that minimizes the impact of measurement error.

To estimate the health effects of simultaneous exposure to multiple pollutants, in this paper we will allow for flexible specification of the air pollution-health outcome risk surface by incorporating interactions among pollutants and allowing for smooth nonlinear functions of pollutant concentrations. For the full BHM, we define $\boldsymbol{\beta}_i$ to be the random effects describing the association between the health outcome and the multiple exposure variables included in the regression model (e.g. nonlinear functions of pollution variables and potential confounders) for the ith location. The parameter of primary scientific interest (θ_i) is the increased health risk when ambient levels of the pollutants considered are simultaneously above their national standards compared to when levels are below their national standards. Our goals are to obtain more precise estimates of θ_i by borrowing strength across locations, estimating overall regional or national risks $\theta^* = \mathbb{E}(\theta_i)$, and identifying site-specific factors (e.g. population demographics, traffic patterns) that modify the association between simultaneous exposure to multiple pollutants and adverse health outcomes.

More generally, the hierarchical modeling approaches we consider apply to problems where the parameter of interest θ_i can be defined as a known function of $\boldsymbol{\beta}_i$ where $\dim(\boldsymbol{\beta}_i) \gg \dim(\theta_i)$. Many difficulties may arise upon implementation of standard Generalized Linear Mixed Models (GLMM) or full BHM in presence of a high-dimensional vector of random effects ($\boldsymbol{\beta}_i$). First, one must specify a multivariate distribution on the full $\boldsymbol{\beta}_i$, which might not be of primary scientific interest. There is an extensive literature on the consequences of misspecification of random-effect distributions in GLMM (Verbeke and Lesaffre, 1997; Heagerty and Kurland, 2001; Litière et al., 2008; Agresti et al., 2004). Though small to moderate misspecification of the random-effect

Hosted by The Berkeley Electronic Press

distribution may not have a large impact in the estimation of fixed effects, there are situations for which misspecification can result in efficiency loss and biased estimates of the random effects (Neuhaus et al., 1992; Heagerty and Kurland, 2001; Agresti et al., 2004; Litière et al., 2010; McCulloch and Neuhaus, 2011). Several approaches have been proposed for specifying flexible semi- or nonparametric distributions for the random effects (Laird, 1978; Magder and Zeger, 1996; Komárek and Lesaffre, 2008; Gallant and Nychka, 1987; Chen et al., 2002). However, most of these approaches cannot be implemented in the context of a high-dimensional vector of random effects, and the validity of the assumption on the random-effect distribution is sometimes difficult to verify (Agresti et al., 2004; Litière et al., 2008). Second, if one is interested in estimating effect modification, then a standard BHM presents the additional challenge of specifying a high-dimensional multivariate regression model. Third, implementing diagnostic methods for misspecification of a multivariate random-effect distribution can be very challenging. Fourth, it may be computationally intensive and/or challenging to implement an MCMC sampler that mixes well and converges quickly to the stationary distribution as the dimension of the vector of random effects increases.

In this paper, we introduce *reduced Bayesian hierarchical models* as a general statistical approach for eliminating nuisance parameters in hierarchical models with a large number of random effects. For reduced BHM we combine information across clusters (e.g. locations) directly on the parameter of interest θ_i. At the first stage, we calculate an integrated likelihood for θ_i, and at the second stage, we specify flexible random-effect distributions directly on the θ_i. Reduced BHM overcome many of the practical challenges in the specification and implementation of full BHM in the context of a high-dimensional vector of nuisance parameters $\boldsymbol{\beta}_i$. Further, we conduct simulation studies to determine the circumstances under which reduced BHM may be more efficient than full BHM. Though developed to study health effects of simultaneous exposure to multiple pollutants, reduced BHM are widely applicable for other studies of multiple exposures, and in general to clustered datasets with a large number of nuisance parameters. Accordingly, the simulation study and much of the methods section are presented in a general context while maintaining a close connection to the scientific motivation for this work.

In Section 2, we describe the multisite time series data used to estimate the health risks associated with simultaneous exposure to multiple pollutants. In Section 3, we describe the level-one model of a BHM aimed at estimating the association between joint exposure to ozone and fine particulate matter and hospital admissions. In Section 4, we introduce the reduced BHM in a general setting where an integrated likelihood is estimated for each cluster and a flexible random-effect distribution is specified directly on the cluster-specific parameter of interest. Section 5 describes our simulation study. In Section 6, we present our results from the data analysis. We provide discussion and concluding remarks in Section 7.

2 Data

We use data from a national database consisting of parallel time series from 60 counties in the northeastern United States during the period 1999-2005. Daily counts of emergency cardiovascular (CVD) hospital admissions (comprising heart failure, heart rhythm disturbances, cerebrovascular events, ischemic heart disease, and peripheral vascular disease), were obtained from billing claims of US Medicare enrollees. CVD admissions were stratified by two age categories, 65–74

COBRA
Collection of Biostatistics
Berkeley Electronic Press

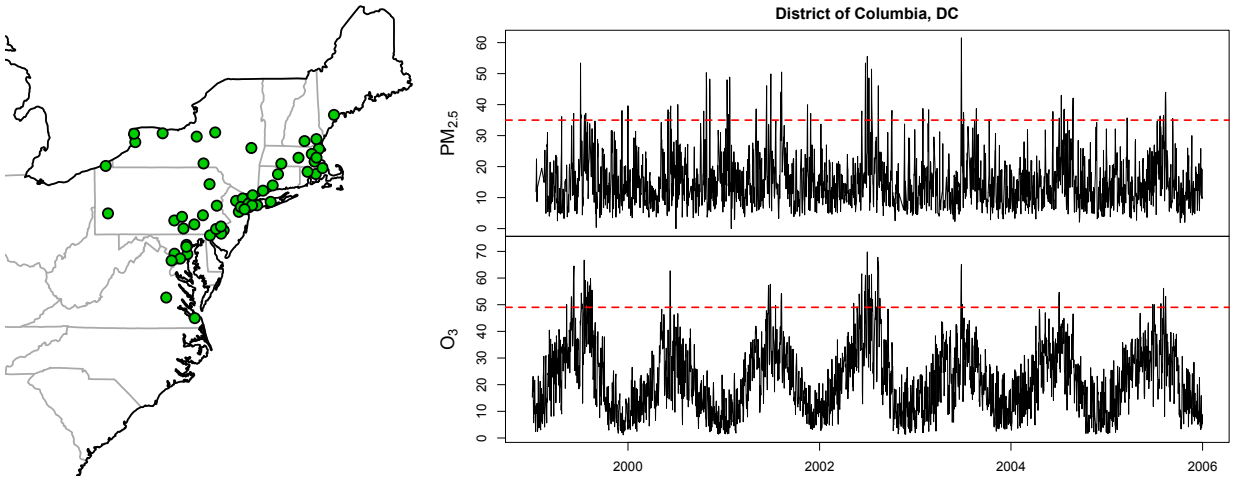

Figure 1: Map of the 51 northeastern US counties. Right panel shows daily times series of $PM_{2.5}$ and O_3 for the District of Columbia. Horizontal line corresponds to the daily national standard for each pollutant.

and $\geqslant 75$. Concentrations of fine particulate matter ($PM_{2.5}$) and ozone (O_3), which for many counties are measured on either a 1-in-3 or 1-in-6 day schedule, were obtained from the US EPA's Air Quality System. Daily temperature and dewpoint temperature were obtained from the National Climatic Data Center. Among the 60 northeastern US counties with available data, we considered the 51 counties having at least 100 days where $PM_{2.5}$ and O_3 were measured concurrently, as well as at least one day when both pollutants were above their national standard (defined below). Figure 1 shows a map of the locations, as well as an example time series of $PM_{2.5}$ and O_3 for Washington, DC.

3 Poisson regression model for multiple pollutants

In this section we describe the first level of a Bayesian hierarchical model for estimating health effects associated with simultaneous exposure to multiple pollutants. We assume for county i on day j for age group k, the number of CVD admissions y_{ijk} has a Poisson distribution with mean model

$$\begin{aligned}\log \mathbb{E}[y_{ijk}] = &\log(n_{ijk}) + \gamma_{i0} + ns(\mathsf{PM}_{2.5ij}; 3\ df, \mathbf{b}_{i1}) \cdot ns(\mathsf{O}_{3ij}; 3\ df, \mathbf{b}_{i2}) \\ &+ \gamma_{i1}\mathtt{age}_k + \boldsymbol{\gamma}'_{i2}\mathtt{dow}_{ij} + ns(\mathtt{temp}_{ij}; 6\ df, \boldsymbol{\gamma}_{i3}) + ns(\mathtt{dptp}_{ij}; 3\ df, \boldsymbol{\gamma}_{i4}) \\ &+ ns(\overline{\mathtt{temp}}^{(3)}_{ij}; 6\ df, \boldsymbol{\gamma}_{i5}) + ns(\overline{\mathtt{dptp}}^{(3)}_{ij}; 3\ df, \boldsymbol{\gamma}_{i6}) + ns(j; 8\ df/\mathsf{year}, \boldsymbol{\gamma}_{i7}),\end{aligned} \tag{1}$$

where n_{ijk} is the number of individuals of the kth age group at risk, and $ns(\cdot; df, \mathbf{b})$ denotes natural cubic splines with the specified degrees of freedom (df). The product of the cubic spline bases for fine particulate matter and ozone, which includes both main effects and interaction terms, provides a flexible specification of the unknown joint pollutant-hospital admissions exposure-response surface. Here $\mathsf{PM}_{2.5ij}$ and O_{3ij} correspond to daily trimmed means (over monitors in county i)

Hosted by The Berkeley Electronic Press

of fine particulate matter and ozone concentrations, respectively; age denotes an indicator for being in the $\geqslant 75$ age category (versus 65 to 74); dow is a vector of indicator variables for day of week; $\texttt{temp}_{ij}$ ($\overline{\texttt{temp}}_{ij}^{(3)}$) is the current day's (average of the previous three days') average temperature; and $\texttt{dptp}_{ij}$ ($\overline{\texttt{dptp}}_{ij}^{(3)}$) is the current day's (average of the previous three days') average dew point temperature. The smooth function of calendar time $ns(j; 8\ df/\text{year}, \boldsymbol{\gamma}_{i7})$ accounts for seasonality and longer-term, time-varying trends in hospital admissions.

To place the within-county model (1) within the more general context of BHM for two-level clustered data, we introduce some notation. Let $\mathbf{b}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2})$ be the vector of random effects for the exposure-response surface characterizing the relation between joint exposure to ozone and fine particulate matter and the health outcome. Let $\boldsymbol{\gamma}_i = (\gamma_{i0}, \boldsymbol{\gamma}_{i1}, \ldots, \boldsymbol{\gamma}_{i7})$ be the vector of random effects describing the association between the confounders and the health outcome, and define $\boldsymbol{\beta}_i = (\mathbf{b}_i, \boldsymbol{\gamma}_i)$. Let $\mathbf{x}_{ij}$ denote the full vector of covariate data for day j in county i, and let $\mathbf{x}_{ij}^{\mathbf{b}}$ denote the 15-dimensional subvector of $\mathbf{x}_{ij}$ that is the concatenation of the basis terms for the main effects and interactions of the spline bases for ozone and fine particulate matter $ns(\text{PM}_{2.5ij}; 3\ df, \mathbf{b}_{i1}) \cdot ns(\text{O}_{3ij}; 3\ df, \mathbf{b}_{i2})$.

We next define a daily indicator variable that summarizes whether a county is in attainment of the national standards jointly for $\text{PM}_{2.5}$ and O_3:

$$\texttt{NAAQS}_{ij} = \begin{cases} 3 & \text{if } \text{PM}_{2.5} > 35\ \mu g/m^3 \text{ and } \text{O}_3 > 0.049 \text{ ppm} \\ 2 & \text{if } \text{PM}_{2.5} > 35\ \mu g/m^3 \text{ and } \text{O}_3 \leqslant 0.049 \text{ ppm} \\ 1 & \text{if } \text{PM}_{2.5} \leqslant 35\ \mu g/m^3 \text{ and } \text{O}_3 > 0.049 \text{ ppm} \\ 0 & \text{if } \text{PM}_{2.5} \leqslant 35\ \mu g/m^3 \text{ and } \text{O}_3 \leqslant 0.049 \text{ ppm}. \end{cases}$$

The values 35 $\mu g/m^3$ and 0.049 ppm were derived from the National Ambient Air Quality Standards (NAAQS), which are defined in Appendix A of the Supplementary Materials.

The parameter of interest θ_i, defined as the log of the average number of CVD admissions on days when both ozone and fine particulate matter are above their respective national standards divided by the average number of CVD admissions on days when both are lower than their respective daily standard adjusted for the potential confounding variables, is given by

$$\theta_i := h(\boldsymbol{\beta}_i; \mathbf{x}_i) = \log \frac{\frac{1}{N_{i3}} \sum_{j:\texttt{NAAQS}_{ij}=3} \exp(\mathbf{b}_i' \mathbf{x}_{ij}^{\mathbf{b}})}{\frac{1}{N_{i0}} \sum_{j:\texttt{NAAQS}_{ij}=0} \exp(\mathbf{b}_i' \mathbf{x}_{ij}^{\mathbf{b}})}, \tag{2}$$

where $\mathbf{b}_i$ is the subvector of $\boldsymbol{\beta}_i$ parameterizing the nonlinear association between O_3 and $\text{PM}_{2.5}$ and the health outcome. Here N_{i3} (N_{i0}) are the number of days when both pollutants are above (below) their national standard in county i during the study period 1999–2005. Derivation of the formulation for the parameter of interest is in Appendix B of the Supplementary Materials.

4 Reduced Bayesian hierarchical model

Rather than specify a full BHM on the large number of random effects $\boldsymbol{\beta}_i$, we define a *reduced BHM* directly on the parameter of interest θ_i:

$$\begin{aligned} \mathbf{y}_i \mid \theta_i &\sim L_i(\theta_i); \quad \text{independent, } i = 1, \ldots, I \\ \theta_i \mid \boldsymbol{\alpha} &\sim \text{RE}(\theta_i \mid \boldsymbol{\alpha}); \quad \text{independent, } i = 1, \ldots, I, \end{aligned} \tag{3}$$

COBRA Collection of Biostatistics Research Archive

http://biostats.bepress.com/jhubiostat/paper230

where $L_i(\theta_i)$ denotes a likelihood function (detailed below) and $\text{RE}(\theta_i \mid \boldsymbol{\alpha})$ denotes an arbitrary random-effect distribution. Note that the likelihood function in general depends on the vector of outcome data from the ith cluster $\mathbf{y}_i$ and on the set of covariate data $\mathbf{x}_i$, though we suppress this dependency in our notation. A prior distribution is placed on $\boldsymbol{\alpha}$.

The reduced BHM may be further generalized by allowing the random-effect distribution $\text{RE}(\theta_i \mid \boldsymbol{\alpha})$ to depend on cluster-level covariates $\mathbf{z}_i$, in order to study potential effect modification. In particular, for a scalar effect modifier the second-stage model may be written as $\theta_i \mid \boldsymbol{\alpha} = \alpha_{0i} + \alpha_1 z_i$ and the random-effect distribution placed on the α_{0i}.

4.1 Integrated Likelihood

In the general setting where the parameter of interest θ_i is a complicated function of the level 1 parameters $\boldsymbol{\beta}_i$ as in (2), we propose to use an integrated likelihood for $L_i(\theta_i)$, which we denote by $L_i^{\int}(\theta)$. For notational simplicity the cluster-specific subscript i is suppressed in what follows. An integrated likelihood for the ith cluster may be expressed as

$$L^{\int}(\theta) \propto f_{\theta|\mathbf{Y}}(\theta \mid \mathbf{y})/\pi_\theta(\theta), \tag{4}$$

where $\pi_\theta(\theta)$ is the prior distribution for θ and $f_{\theta|\mathbf{Y}}$ is the corresponding posterior distribution of θ based on the data from only that cluster. Note that in the special case where the cluster-specific parameters $\boldsymbol{\beta}$ can be reparameterized as $(\theta, \boldsymbol{\lambda})$, this expression can be rewritten as $L^{\int}(\theta) = \int L(\theta, \boldsymbol{\lambda})\pi(\boldsymbol{\lambda} \mid \theta)d\boldsymbol{\lambda}$, where $L(\theta, \boldsymbol{\lambda})$ is the joint likelihood, and $\pi(\boldsymbol{\lambda} \mid \theta)$ is the prior density of $\boldsymbol{\lambda}$ given θ (Berger et al., 1999).

When $L^{\int}(\theta)$ is not available in closed form, we propose a simulation approach to approximate (4) as follows:

1. Assign a vague prior distribution to the vector $\boldsymbol{\beta}$ of level 1 parameters, which induces a prior distribution on $\theta = h(\boldsymbol{\beta}; \mathbf{x})$. Simulate R prior samples from $\pi_\theta(\theta)$.
2. Fit a within-cluster Bayesian model to generate R samples $\boldsymbol{\beta}^{(r)}$ from $f_{\boldsymbol{\beta}|\mathbf{y}}(\boldsymbol{\beta} \mid \mathbf{y})$.
3. Obtain the posterior samples $\theta^{(r)} = h(\boldsymbol{\beta}^{(r)}; \mathbf{x})$.
4. Apply a kernel density smoother to estimate $f_{\theta|\mathbf{y}}(\theta \mid \mathbf{y})$ and $\pi_\theta(\theta)$ on the same grid of points.

We repeat this process for each cluster i to obtain approximations $\hat{L}_i^{\int}(\theta_i)$, $i = 1, \ldots, I$. Since this step is performed a single time prior to fitting the reduced BHM, estimating the parameters of the reduced BHM remains fast. Further details of our implementation are in Appendix C of the Supplementary Materials.

4.2 Dirichlet process mixture model for RE$(\theta_i \mid \boldsymbol{\alpha})$

To allow for flexible specification of the random-effect distribution we propose to use a Dirichlet process mixture model for $\text{RE}(\theta_i \mid \boldsymbol{\alpha})$. The Dirichlet process mixture model (Ferguson, 1973;

CORA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Neal, 2000) can be expressed as the limit as the number of components K goes to infinity of

$$\begin{aligned}
\theta_i \mid c_i, \boldsymbol{\phi} &\sim F(\theta_i \mid \boldsymbol{\phi}_{c_i}); \quad \text{independent, } i = 1, \ldots, I \\
c_i \mid \mathbf{p} &\sim \text{Discrete}(p_1, \ldots, p_K); \quad \text{independent, } i = 1, \ldots, I \\
\phi_c &\sim G_0 \\
\mathbf{p} &\sim \text{Dirichlet}(\alpha/K, \ldots, \alpha/K),
\end{aligned}$$

where α/K is the concentration parameter written so that it approaches 0 as K goes to infinity. Here we consider a normal mixture so that $F(\cdot \mid \boldsymbol{\phi}_c) = \text{N}(\cdot \mid \mu_c, \tau_c)$, and we select the conjugate prior so that $G_0 = \text{NormalGamma}(\lambda, \gamma, a, b)$, i.e. $\tau_c \sim \text{Gamma}(\tau \mid a, b)$ and $\mu_c \mid \tau_c \sim \text{N}(\lambda, \gamma\tau_c)$.

4.3 Computational details

The reduced BHM (3) may be fit using Markov Chain Monte Carlo (MCMC) methods to generate samples from the posterior distribution of the unknown parameters

$$\mathbb{P}(\theta_1, \ldots, \theta_I, \boldsymbol{\alpha} \mid \mathbf{y}_1, \ldots, \mathbf{y}_I) \propto \pi(\boldsymbol{\alpha}) \prod_{i=1}^{I} \{\text{RE}(\theta_i \mid \boldsymbol{\alpha}) L_i(\theta_i)\},$$

where $\pi(\boldsymbol{\alpha})$ denotes the prior distribution on the vector of parameters of the random-effect distribution. At each iteration of the MCMC algorithm, a sample is drawn from the full conditional

$$f_c(\theta_i) \propto \text{RE}(\theta_i \mid \boldsymbol{\alpha}) L_i(\theta_i) \tag{5}$$

for each cluster i. When using an estimated integrated likelihood, we replace $L_i(\theta_i)$ in equation (5) by $\hat{L}_i^{\int}(\theta_i)$. Since $f_c(\theta_i)$ is not a known distribution, we sample from it by applying a Metropolis-Hastings step. In the Metropolis-Hastings step, we need to evaluate the likelihood $\hat{L}_i^{\int}$ at an arbitrary point x. We do this by selecting the grid point t_k that is closest to x and evaluating the likelihood $\hat{L}_i^{\int}(x)$ at that grid point.

For generating posterior samples of $\boldsymbol{\alpha}$ when $\text{RE}(\theta_i \mid \boldsymbol{\alpha})$ is the Dirichlet process mixture model defined in Section 4.2, we adapt an MCMC sampling algorithm described by Neal (2000). Details are in Appendix C of the Supplementary Materials.

5 Simulation study

There are instances for which the reduced BHM may be preferred to the full BHM due to practical considerations such as its simplified implementation and the ease with which prior information may be incorporated directly on the parameter of interest. However, a more thorough understanding of situations when the reduced BHM works well is needed. In this section we conduct simulation studies to compare performance of the reduced BHM to the full BHM across a range of scenarios.

We base our studies on data from a meta-analysis of 41 randomized trials of a treatment for stomach ulcers, provided by Efron (1996). The data from the ith experiment is $\{\mathbf{y}_i = (y_{i0}, y_{i1}), \mathbf{x}_i = (n_{i0}, n_{i1})\}$, where y_{i0}, y_{i1} are the number of occurrences of ulcers for the control and treatment groups, and n_{i0}, n_{i1} are the number of subjects in the control and treatment

http://biostats.bepress.com/jhubiostat/paper230

groups, respectively. Let $\mathbf{p}_i = (p_{i0}, p_{i1})$ be the vector of probabilities of the occurrence of ulcers in the control and treatment groups. The distribution of the data from experiment (cluster) i is assumed to be $\mathbb{P}_i(\mathbf{y}_i \mid \mathbf{x}_i; \mathbf{p}_i) = \binom{n_{i1}}{y_{i1}} p_{i1}^{y_{i1}} (1-p_{i1})^{n_{i1}-y_{i1}} \binom{n_{i0}}{y_{i0}} p_{i0}^{y_{i0}} (1-p_{i0})^{n_{i0}-y_{i0}}$, and the parameter of interest is the log odds ratio

$$\theta_i = h(\mathbf{p}_i) = \log \frac{p_{i1}/(1-p_{i1})}{p_{i0}/(1-p_{i0})}. \tag{6}$$

In this example, a full BHM would require the specification of a random-effect distribution for $\mathbf{p}_i = (p_{i1}, p_{i0})$. Alternatively, a commonly used specification first defines a one-to-one transformation of the $\mathbf{p}_i$ into $\mathbb{R}^2$ through the logit link and assumes a bivariate normal distribution for the random effects:

$$\begin{aligned} y_{ki} \mid p_{ik} &\sim \text{Binom}(n_{ik}, p_{ik}) \quad \text{for } k = 0, 1 \\ \text{logit}(p_{ik}) &= \beta_{i0} + \beta_{i1}\mathbf{I}(k=1) \\ (\beta_{i0}, \beta_{i1})' &\sim \text{N}((\beta_0^*, \beta_1^*)', \Sigma). \end{aligned} \tag{7}$$

For a reduced BHM, one first summarizes the information contained in experiment i about the log odds ratio θ_i through a likelihood function, and then specifies a random-effect distribution directly on the θ_i. For this problem, a conditional likelihood for θ_i is available in closed form. By conditioning on the margins of the two-by-two table for each experiment, the conditional likelihood may be expressed as

$$L_i^C(\theta_i) = \frac{\binom{n_{i0}}{y_{i0}} \binom{n_{i1}}{y_{i1}} \exp(\theta_i y_{i1})}{\sum_{u=0}^{\min(n_{i1}, y_{i0}+y_{i1})} \binom{n_{i0}}{u} \binom{n_{i1}}{y_{i1}+y_{i0}-u} \exp(\theta_i u)}. \tag{8}$$

We may then use $L_i^C(\theta_i)$ for the likelihood function in the reduced BHM (3).

In our simulation study we will simulate data under four data generating mechanisms, and we will estimate model parameters under four BHM formulations. First we will describe each of the hierarchical modeling approaches used to fit the data, and then we will detail the four data generating models.

5.1 Bayesian hierarchical models

Each simulated data set was fit using four approaches: a full BHM assuming the logistic model (7) with a normal random-effect distribution on the $\boldsymbol{\beta}_i$ (FBHM); a reduced BHM using the conditional likelihood $L_i^c(\theta_i)$ from equation (8) with a normal random-effect distribution on the θ_i (RBHM-L-N); a reduced BHM using the conditional likelihood $L_i^c(\theta_i)$ from (8) with a flexible random-effect distribution on the θ_i (RBHM-L-DP); and a reduced BHM using a normal approximation to the likelihood with a normal random-effect distribution on the θ_i (RBHM-N-N). For the flexible random-effect distribution, we considered the Dirichelet Process normal mixture model described in Section 4.2. For the reduced BHM, since a conditional likelihood for θ_i is available in closed form, we do not consider an integrated likelihood as the likelihood function for this study. For each approach, we estimate the cluster-specific log odds ratios θ_i as well as the overall log odds ratio $\theta^* = \mathbb{E}(\theta_i)$. Additionally we obtain 95% posterior intervals for the overall and cluster-specific parameters. Details of estimation for each of the four models are in Appendix D of the Supplementary Materials.

Hosted by The Berkeley Electronic Press

5.2 Data generating models

We consider four data generating models. We always assume $y_{i0} \sim \text{Binom}(n_{i0}, p_{i0})$ and $y_{i1} \sim \text{Binom}(n_{i1}, p_{i1})$, and we select different models for generating p_{i0} and p_{i1} $(i = 1, \ldots, I)$. Note that each model for generating p_{i0} and p_{i1} induces a distribution on the log odds ratio θ_i through (6). Thus, each time we generate a dataset, we obtain I values of the cluster-specific, true log odds ratios θ_i (one for each cluster i). The models were selected in order to distinguish among scenarios where the full BHM is expected to outperform the reduced BHM and vice versa. Figure 2 shows,

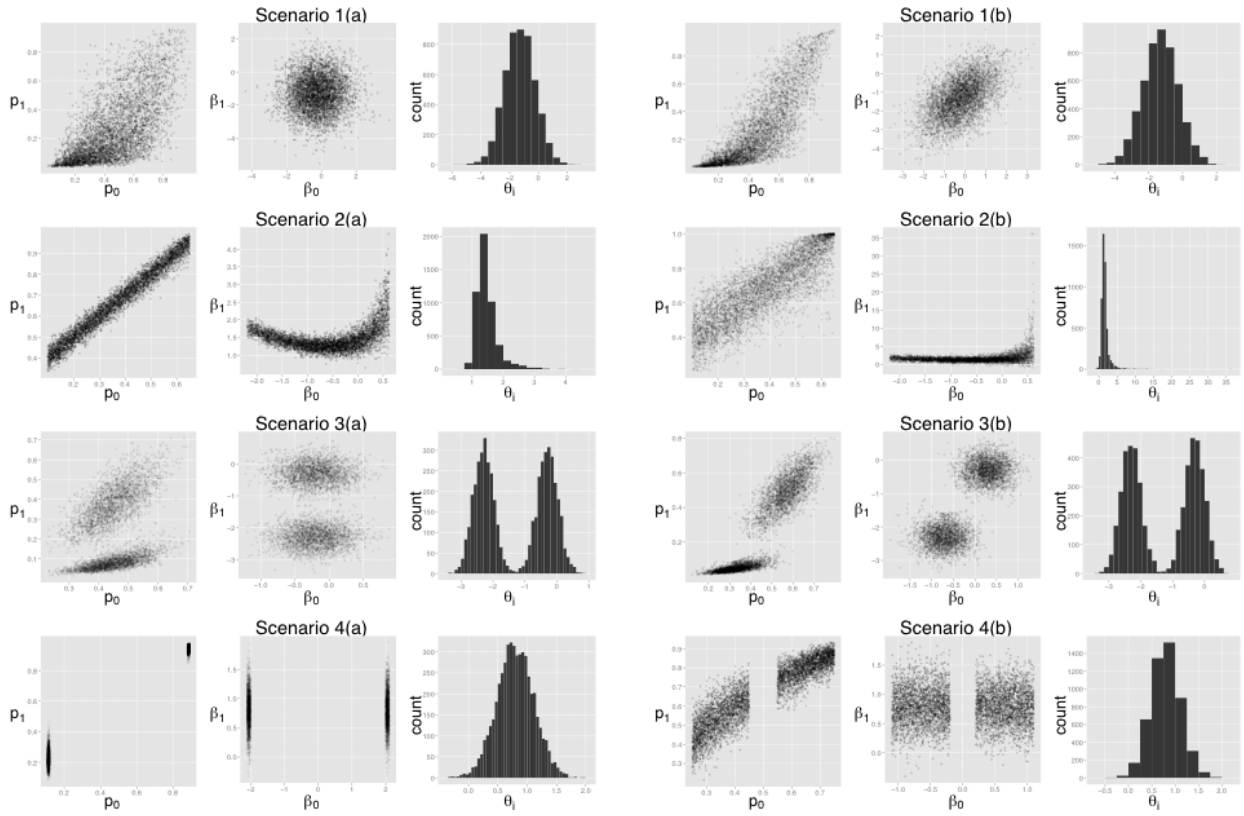

Figure 2: Plots of simulated data under each scenario from the four models. First row displays data from model 1, scenarios (a)–(b); row 2 shows data from model 2, scenarios (a)–(b); row 3 corresponds to model 3, scenarios (a)–(b); and row 4 to model 4, scenarios (a)–(b). For each scenario 5000 data points (p_{i0}, p_{i1}) are plotted, as well as the corresponding points (β_{i0}, β_{i1}) under the transformation $\text{logit}(p_{ik}) = \beta_{i0} + \beta_{i1}\mathbf{I}(k = 1)$, and histograms of the corresponding log odds ratios θ_i.

for each of the four data generating models, the distribution of the (p_{i0}, p_{i1}), along with the corresponding distributions of the $(\beta_{0i}, \beta_{1i}) = (\log \frac{p_{i0}}{1-p_{i0}}, \theta_i)$ and the log odds ratios θ_i.

In each case, we set $n_{i0} = n_{i1} = n$, and we considered $n = 100$ for both $I = 100$ and $I = 50$. These parameter values were selected to correspond to a large within-cluster sample size for either a large or moderate number of clusters.

Collection of Biostatistics Research Archive

Model 1 - Bivariate Normal
We generate data from

$$(\beta_{0i}, \beta_{1i})' \sim \mathsf{N}((\beta_0^*, \beta_1^*)', \Sigma)$$
$$\mathsf{logit}(p_{ki}) = \beta_{0i} + \beta_{1i}\mathbf{I}(k = 1),$$

where $(\beta_0^*, \beta_1^*) = (-0.2, -1.3)$, and we consider two different values for Σ,

$$\Sigma_a = \begin{bmatrix} 0.9 & 0 \\ 0 & 1.1 \end{bmatrix} \text{ and } \Sigma_b = \begin{bmatrix} 0.9 & 0.5 \\ 0.5 & 1.1 \end{bmatrix}.$$

These parameter values were selected to be the same order of magnitude of those from the ulcer data set. Since this model fully specifies a normal random-effect distribution on the $\boldsymbol{\beta}_i$, particularly in scenario 1(b) where a moderate correlation between the random effects is assumed, we expect it to favor the full BHM (7).

Model 2 - Uniform/Beta
We generate $p_{i0} \sim \mathsf{Uniform}(0.1, 0.6)$ and $p_{i1} \mid p_{i0} \sim \mathsf{Beta}(m = p_{i0} + 0.3, \phi)$. Here we have parameterized the beta distribution by its mean m and variance ϕ. We consider two values for ϕ, namely $\phi_a = 0.001$ and $\phi_b = 0.01$. Since this model is not based on either the full or reduced BHM *a priori* we don't expect it to favor either of these two approaches.

Model 3 - Normal Mixture
We generate (p_{i0}, p_{i1}) by

$$(\beta_{0i}, \beta_{1i})' \sim \alpha\mathsf{N}(\boldsymbol{\beta}^* - \boldsymbol{\nu}, \Sigma) + (1 - \alpha)\mathsf{N}(\boldsymbol{\beta}^* + \boldsymbol{\nu}, \Sigma)$$
$$\mathsf{logit}(p_{ki}) = \beta_{0i} + \beta_{1i}\mathbf{I}(k = 1),$$

where we fix $\boldsymbol{\beta}^* = (-0.2, 1.3)$, $\alpha = 0.5$, and $\Sigma = \mathsf{diag}\{(0.01, 0.01)\}$. We consider two values for $\boldsymbol{\nu}$, namely $\boldsymbol{\nu}_a' = (0, 1)$ and $\boldsymbol{\nu}_b' = (0.5, 1)$. This data generating model was selected because the random-effect distribution will be misspecified for both the full and the reduced BHM (since $\theta_i = \beta_{i1}$), when a normal random-effect distribution is assumed; thus, we expect neither approach to perform particularly well.

Model 4 - Normal-θ_i
Finally, we generate data by first simulating values for the log odds ratios θ_i and for the log odds $\lambda_i = \log(\frac{p_{0i}}{1-p_{0i}})$, which induces a distribution on the $(p_{0i}, p_{1i}) = \left(\frac{\exp(\lambda_i)}{1+\exp(\lambda_i)}, \frac{\exp(\lambda_i+\theta_i)}{1+\exp(\lambda_i+\theta_i)}\right)$. In particular, we simulate $\theta_i \sim \mathsf{N}(\mu, \sigma^2)$ and $\lambda_i \sim 0.5\mathsf{U}(-u_2, -u_1) + 0.5\mathsf{U}(u_1, u_2)$, where we fix $\mu = 0.8$, $\sigma^2 = 10$. We consider two scenarios for u_1 and u_2, namely $(u_{1a}, u_{2a}) = (2, 2.1)$ and $(u_{1b}, u_{2b}) = (0.2, 1.1)$. This model was chosen because it is expected to favor the reduced BHM over the full BHM, since the normal random-effect distribution on the $(\beta_{i0}, \beta_{i1})'$ for the full BHM will be misspecified, while the random-effect distribution for θ_i in the reduced BHM will be correctly specified.

5.3 Results

We compare the posterior mean estimates $\tilde{\theta}_i$ of the cluster-specific log odds ratios from the four modeling approaches (FBHM, RBHM-L-N, RBHM-L-DP, and RBHM-N-N) to their true

Hosted by The Berkeley Electronic Press

generated values θ_i using as risk criterion the squared error loss $\sum_{i=1}^{I}(\tilde{\theta}_i - \theta_i)^2$. We also evaluate the estimators $\tilde{\theta}^*$ for the overall $\theta^* = \mathbb{E}(\theta_i)$, by comparing the root mean squared error (rMSE), bias, and variance from each approach. In addition, we evaluate the coverage of 95% posterior intervals for $\tilde{\theta}_i$ and $\tilde{\theta}^*$. Table 1 displays results across the different methods for the case $I = 100$. Results for the case $I = 50$ are in Table 1 of the Supplementary Materials.

Overall, we found that the reduced BHM (RBHM-L-N and RBHM-L-DP) performed comparably to the full BHM across the majority of scenarios and even had superior performance in some cases. Specifically, except under data generating models 1(b) and 3(b), we found that RBHM-L-DP performed as well as or better than FBHM for estimating the θ_i. All of the approaches (except RBHM-N-N) performed comparably for estimating θ^*.

The main disparity in performance across these methods occurred for estimation of the cluster-specific parameters θ_i. Situations for which FBHM is only marginally better or worse than the reduced BHM are those for which a bivariate normal distribution fails to capture the relationship between β_{0i} and β_{1i}, as for data generating models 2, 3(a), and 4. The two situations where FBHM yielded the best cluster-specific estimates were ones where the data generating model implies considerable correlation in β_{0i} and β_{1i}, which can be captured to varying degrees by the bivariate normal random-effect distribution on the $\boldsymbol{\beta}_i$. This occurred for data generating models 1(b) and 3(b), which had correlation of $\approx$ 0.5 and 0.8, respectively (see Figure 2). Because nuisance parameters are eliminated before pooling, the reduced BHM do not take advantage of this correlation structure. Comparing the reduced BHM with different random-effect distributions, we found that RBHM-L-DP performed just as well or only slightly worse than RBHM-L-N when the true distribution was normal (models 1(a)–(b) and 4(a)–(b)), but performed moderately better when the true random-effect distribution was non-normal (models 2(a)–(b) and 3(a)–(b)).

Across simulation scenarios we generally found that the model using the normal approximation to the likelihood (RBHM-N-N), although most efficient computationally, was not competitive with the other approaches. For estimating θ_i, the RBHM-N-N either performed comparably (scenarios 2(a), 3(a) and 4(a)–(b)), or moderately worse (scenarios 1(a)–(b), 2(b), and 3(b)) than the other approaches. For estimating the overall θ^*, the RBHM-N-N generally had larger rMSE and coverage markedly lower than the nominal rate (exceptions are scenarios 3(a) and 4(c)). One reason for the poor performance of RBHM-N-N is that the normal approximation to the likelihood does not provide a good approximation in this application, particularly when y_{i1} or y_{i0} is equal to zero or n.

6 Application

We applied the reduced BHM defined in Section 4 to our multisite time series study of 51 urban counties in the northeastern US for the period 1999–2005. Our goal was to estimate the county-specific and overall log relative risks of emergency cardiovascular hospital admissions associated with levels of $PM_{2.5}$ and O_3 above their national standards.

We considered three types of reduced BHM. The first uses a normal approximation to the likelihood at the first stage and a normal distribution on the random effects at the second stage (RBHM-N-N). The second uses an integrated likelihood at the first stage and a normal random-effect distribution at the second stage (RBHM-L-N). The third uses an integrated likelihood at the first stage and a Dirichlet process normal mixture for the random-effect distribution (RBHM-

Collection of Biostatistics Research Archive

L-DP). The parameter of interest θ_i, defined in (2), is the log relative risk of cardiovascular admissions when ozone and $PM_{2.5}$ are both above their national standarda compared to when both are in attainment. For each reduced BHM we assumed little prior information, by incorporating diffuse priors on the overall θ^*. We first fit each reduced BHM without including any second-level covariates. We subsequently considered inclusion, at the second stage, of a county-specific measure of the average level of NO_2 during the study period to demonstrate the identification of effect modification using reduced BHM. Long-term average NO_2 may be an important effect modifier because it a proxy for traffic exposure. This was done by assuming, at the second level that $\theta_i = \alpha_{0i} + \alpha_1 z_i$, where z_i is the long-term average NO_2 for the ith county, and placing a random-effect distribution on the α_{0i}. Details of the implementations for each reduced BHM are in Appendix C of the Supplementary Materials.

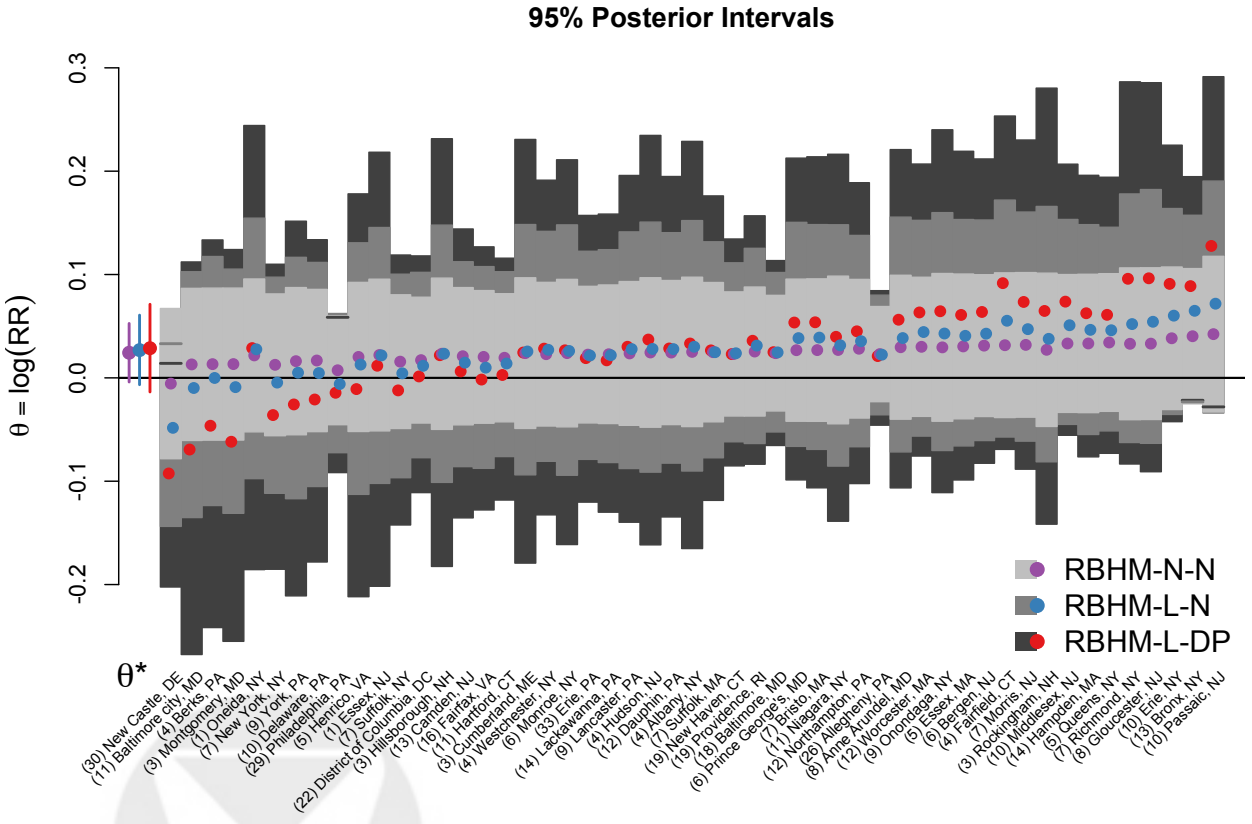

Figure 3: For 51 northeastern US counties, 95% posterior intervals for θ_i and the overall θ^* for each reduced BHM. Models considered are the normal approximation to the likelihood with normal random-effect distribution (RBHM-N-N) and integrated likelihood with normal (RBHM-L-N) and flexible (RBHM-L-DP) random-effect distributions. Counties ordered from left to right by increasing values of $\hat{\theta}_i/\widehat{SD}_i$ where $\hat{\theta}_i$ is the MLE and $\widehat{SD}_i$ is the estimated standard error. Number of days with both O_3 and $PM_{2.5}$ in exceedence of their national standards listed beside each city. The parameter of interest θ is the log relative risk of cardiovascular admissions on days when both O_3 and $PM_{2.5}$ exceed the national standard compared to days when both are in attainment.

Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Figure 3 shows the posterior mean estimates and 95% posterior intervals for the overall θ^* and for the cluster-specific θ_i obtained under each reduced BHM. We found that on average, across all counties, there was an increase in CVD admissions on days when both ozone and fine particulate matter were above their national standards compared to days when both pollutants were in attainment. In particular, we estimated that the overall log relative risk of CVD admissions associated with levels of O_3 and $PM_{2.5}$ both above their national standards (θ^*) was 0.024 (95% posterior interval -0.004 to 0.053) for RBHM-L-N, 0.027 (-0.007 to 0.061) for RBHM-L-N, and 0.029 (-0.014 to 0.071) for RBHM-L-DP. A log relative risk of 0.024 corresponds (approximately) to a 2.4% increase in cardiovascular hospital admissions on days when both O_3 and $PM_{2.5}$ are above their standards compared to when both pollutants are in attainment. We also found variability across counties in the estimate of the cluster-specific effects θ_i. For most counties, θ_i was estimated to be positive, though for each county the posterior interval covered zero. The random-effect estimates exhibited the largest shrinkage for RBHM-N-N, followed by RBHM-L-N, with the RBHM-L-DP estimates remaining furthest from the overall regional estimate.

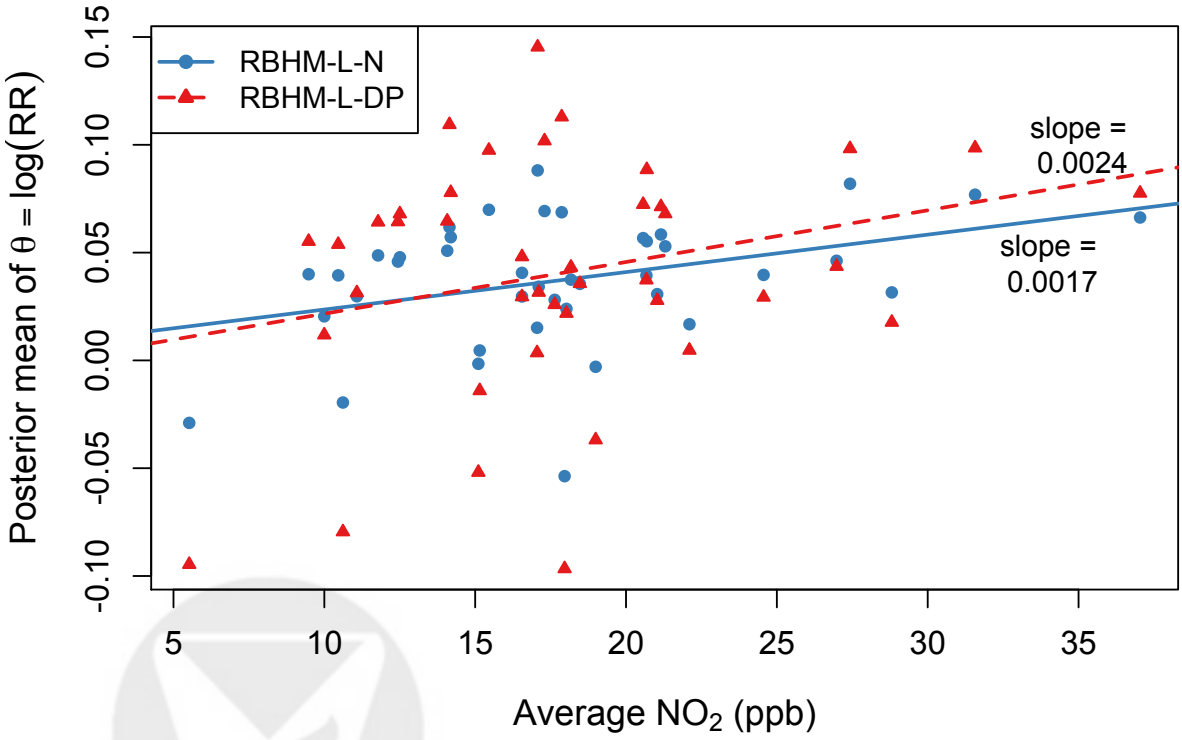

Figure 4: For the 41 northeastern US counties with NO_2 measurements, plot of the posterior mean of θ_i from each county estimated from the reduced BHM including the average annual NO_2 at the second level versus the average annual NO_2. The parameter of interest θ_i is the log relative risk of cardiovascular admissions on days when both O_3 and $PM_{2.5}$ exceed their national standard compared to days when both are in attainment.

Figure 4 shows the posterior mean estimates of the location-specific θ_i from the reduced BHM including average NO_2 as a covariate in the second stage, plotted against the location's long-term average NO_2. The positive slopes (α_1) suggest that the risk of cardiovascular admissions

http://biostats.bepress.com/jhubiostat/paper230

associated with daily levels of O_3 and $PM_{2.5}$ greater than their national standards is higher in locations with greater NO_2 levels and lower in locations with lower NO_2 levels, though the estimates were not statistically significant. More precisely, we estimated that an interquartile range increase in long-term average NO_2 is associated with a percentage increase in the relative risk of cardiovascular hospital admissions associated with O_3 and $PM_{2.5}$ both above their national standards of 1.2% (-3.8% to 6.2%) under RBHM-L-N, and 1.6% (-2.2% to 5.7%) under RBHM-L-DP.

7 Discussion

While previous studies have estimated health effects of single pollutants, understanding how complex mixtures of pollutants affect health remains a challenging goal. Quantifying health risks resulting from exposure to a single pollutant is a useful analytical construct, but it is not representative of true exposure. It is therefore critical to develop models for estimating health effects associated with simultaneous exposure to multiple pollutants.

In this paper we estimate county-specific and regional average multipollutant risks by investigating the joint effect of fine particulate matter and ozone in a two-pollutant exposure-response model. We extend previous single pollutant models by allowing for nonlinear smooth functions of multiple pollutants and their interactions at the first stage of the model and for effect modification at the second stage. Because we model flexible associations of several exposures concurrently, the inclusion of interactions of spline terms leads to a high-dimensional vector of random effects. As a result, several challenges to the application of the usual full BHM framework are introduced. To address these challenges, we propose the *reduced BHM* as a novel approach for combining information across locations directly on the parameter of interest. In this approach, information about the parameter of interest is summarized through a likelihood function (e.g. integrated likelihood) in the first stage. At the second stage, a flexible random-effect distribution (e.g. Dirichlet process mixture) is specified directly on the parameter of interest. We conducted simulation studies to compare performance of the reduced BHM to the full BHM, and we applied the reduced BHM to a multisite time series study of 51 northeastern US counties during the period 1999–2005.

In comparison with the reduced BHM, on first inspection the full BHM is the seemingly optimal approach, as it uses all of the available data in a single model to combine information across clusters. However, many practical difficulties may arise upon implementation. First, for the full BHM one must specify a random-effect distribution on the $\boldsymbol{\beta}_i$, which may be difficult when the $\boldsymbol{\beta}_i$ are high-dimensional or when they do not have meaningful interpretations (e.g. regression spline coefficients as in equation (1)). A multivariate normal distribution is frequently used, though it is not a requirement. For example, more flexible parametric distributions have been proposed for the random-effect distribution, including semiparametric and nonparametric modeling approaches (Magder and Zeger, 1996; Komárek and Lesaffre, 2008; Chen et al., 2002), but these may be challenging to implement in a high-dimensional context. Additionally, prior distributions must be selected for the parameters of the random-effect distribution (e.g. mean vector $\boldsymbol{\beta}^*$ and variance-covariance matrix Σ), which may also be complicated if these parameters do not have meaningful interpretations. If there does not exist a reparameterization of $\boldsymbol{\beta}_i$ such that $\boldsymbol{\beta}_i = (\theta_i, \boldsymbol{\lambda}_i)$ for $\boldsymbol{\lambda}_i$ a $(q-1)$-dimensional nuisance parameter, then prior information about the quantity of interest $\theta_i = h(\boldsymbol{\beta}_i; \mathbf{x}_i)$ cannot be easily translated into prior information about

Hosted by The Berkeley Electronic Press

the model parameters $\boldsymbol{\beta}_i$. In addition, if one is interested in effect modification of cluster-specific covariates $\mathbf{z}_i$ at the second level, then a potentially high-dimensional multivariate regression model for $\boldsymbol{\beta}_i \mid \boldsymbol{\beta}^*, \mathbf{z}_i$ must be specified. Finally, implementation of the MCMC sampler will become increasingly challenging and computationally intensive as the dimension of $\boldsymbol{\beta}_i$ (number of random effects) increases.

For the reduced BHM, on the other hand, rather than specify a high-dimensional random-effect distribution on parameters that are not of primary scientific interest, one only needs to specify a random-effect distribution for a one-dimensional parameter that has a meaningful interpretation. Additionally, it is frequently much easier to incorporate prior information about the parameter of interest θ_i than about a large vector of nuisance parameters $\boldsymbol{\beta}_i$ that may be hard to interpret (e.g. spline coefficients). Furthermore, reducing a hierarchical model on a high-dimensional vector of parameters to a hierarchical model on a much lower dimensional space yields simpler implementation and greater computational efficiency, and makes model diagnostics and sensitivity analyses more wieldy.

While the reduced BHM overcomes many of the difficulties in the specification and implementation of the full BHM, it also introduces new challenges. First, one must deal with the elimination of nuisance parameters in the first step to obtain $L_i(\theta_i)$. While the literature on likelihood-based methods for eliminating nuisance parameters is vast (Pawitan, 2001; Edwards, 1992), in this paper we restricted our attention to those likelihoods that correspond to true probability distributions, including the integrated and conditional likelihood functions. In the case of large within-cluster sample sizes, the choice of which likelihood function to use should make little difference compared with the impact of the selection of the random-effect distribution. Future work could further investigate how different choices of the likelihood function may influence inference in fitting the reduced BHM. Second, while one gains simplicity by eliminating nuisance parameters at the outset, it is possible that some information may be lost before combining information across clusters. For example, if there is large correlation among the components of $\boldsymbol{\beta}_i$ within a cluster, then reducing the parameter space to a single parameter and pooling the θ_i may not fully capture this underlying correlation and may yield less efficient estimates. We investigated this possibility in our simulation study, finding that large correlation of random effects within a cluster generally led to improved estimation of the random effects by the full BHM compared to the reduced BHM. However, in other scenarios, namely those for which the random-effect distribution for the full BHM was misspecified, the reduced BHM achieved superior performance. In addition, for estimating the overall θ^* we found performance to be very similar across methods. Overall, in our simulation studies the reduced BHM performed nearly as well as the full BHM (which uses full information), and even performed better in some cases.

While BHM using likelihoods of a single parameter in the level 1 model have been proposed previously for specific applications, including the meta analysis of randomized trials of a treatment for stomach ulcers (Efron, 1996; Liao, 1999), to our knowledge the reduced BHM has not been described in the very general form we have presented here. In our formulation, the reduced BHM is applicable when there is no conditional or marginal likelihood available, when an integrated likelihood is not available in closed form, when the second level model includes cluster-specific covariates, and when flexible specifications of the random-effect distribution are desired. Further, we have not seen a direct comparison of the reduced BHM to the full BHM across a range of potential scenarios.

Development of the reduced BHM was motivated by methodological needs for estimating

http://biostats.bepress.com/jhubiostat/paper230

health risks of joint exposure to multiple pollutants. We applied the reduced BHM methodology to estimate the risk of emergency cardiovascular admissions associated with simultaneous exposure to fine particulate matter and ozone. For the overall effect θ^*, we found marginal evidence of increased risk on days when both pollutants exceeded their national standards compared to when both were in attainment. The reduced BHM with normal random-effect distribution on the parameter of interest θ_i (RBHM-L-N) led to more shrinkage of the county-specific random effects than the reduced BHM with flexible random-effect distribution (RBHM-L-DP). Further, the RBHM-L-N had narrower credible intervals for the county-specific parameters θ_i than RBHM-L-DP. If the normal random-effect distribution is misspecified (e.g. if the analysis is missing an important county-level effect modifier) then the RBHM-L-N may understate statistical uncertainty in the θ_i. We also demonstrated that the reduced BHM can easily accommodate effect modifiers. Specifically, we examined the inclusion of long-term county-level NO_2, a surrogate for traffic exposure. We found a larger relative risk of cardiovascular admissions associated with levels of $PM_{2.5}$ and O_3 higher than their national standards in locations with high average NO_2 compared to locations with low average NO_2, although the effect modification was not statistically significant. In the future we will apply this approach to systematically conduct a national epidemiological investigation of the health effects associated with simultaneous exposure to multiple pollutants.

We have described the reduced BHM methodology within the context of estimating health risks of exposure to many pollutants. However, this hierarchical modeling strategy is broadly applicable in situations where the parameter of interest is a known function of stage 1 random effects. The meta-analysis of peptic ulcer treatment that served as the basis for our simulation study is one example. Another example is the estimation of heat wave mortality risk in multisite time series studies (Bobb et al., 2011). One can build a location-specific model similar to (1) where the exposure-response function of interest is the temperature-mortality relation, adjusted for time-varying covariates. One can then define a heat wave day indicator variable as a function of temperature on current and previous days. The parameter of interest θ, defined as the log relative risk of mortality on heat wave days compared to non-heat wave days, can then be written as a known function of the temperature-response function (parameterized by $\boldsymbol{\beta}$), and the reduced BHM framework may be applied.

There are several extensions to the reduced BHM methodology we proposed. First, we assumed a within-location model that had the same form across locations. However, this assumption could be relaxed. To implement the reduced BHM, one can specify different within-cluster models for each cluster, as long as the interpretation of the parameter of interest remains constant across models. For example, for the within-cluster model (1) in the multipollutant application, the full BHM would require a common spline basis (e.g. common knot locations) for the joint O_3 and $PM_{2.5}$ association across locations, while the reduced BHM can allow for locally optimized spline bases. Thus the reduced BHM approach can readily accommodate heterogeneity in the appropriate model to use across locations. In this manuscript we focused on two-level clustered datasets and a scalar parameter of interest. However, the reduced BHM could be generalized to three-level or higher-level models, and to situations where the parameter of interest $\boldsymbol{\theta}_i = h(\boldsymbol{\beta}_i)$ is a multivariate parameter with $\dim(\boldsymbol{\theta}_i) < \dim(\boldsymbol{\beta}_i)$.

While developed to study health risks associated with multiple exposures, the proposed reduced BHM methodology is widely applicable to studies with clustered data in which the parameter of interest is a known function of the parameters $\boldsymbol{\beta}_i$ of the within-cluster model. This framework is especially useful when $\boldsymbol{\beta}_i$ is high-dimensional, when the components of $\boldsymbol{\beta}_i$ are not

Hosted by The Berkeley Electronic Press

easily interpretable, or when one wishes to incorporate prior information directly on the parameter of interest. For such applications, the reduced BHM allows one to specify a random-effect distribution directly on the parameter of interest θ_i and to study effect modification by specifying an across-cluster regression model for θ_i. Further, the reduced parameter space leads to simpler implementation, which facilitates the specification of flexible random-effect distributions that do not require strong assumptions on the random effects. For problems that are very high-dimensional in the number of clusters, the number of observations within a cluster, and the number of parameters in the within-cluster model, it may not be computationally feasible to fit a full BHM. In such cases, the reduced BHM is a practical alternative.

Supplementary Materials

Appendices and Tables referenced in Sections 3, 4.1, 4.3, 5.1, 5.3, and 6 are available upon request from the first author.

Acknowledgments

This work was supported by the National Institute of Environmental Health Sciences [grant numbers T32ES012871, R01ES012054, R01ES019560]; and the United States Environmental Protection Agency [grant numbers RD-83479801, RD-83241701].

References

Agresti, A., B. Caffo, and P. Ohman-Strickland (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis 47*, 639–653.

Bell, M., A. McDermott, S. Zeger, J. Samet, and F. Dominici (2004). Ozone and short-term mortality in 95 US urban communities, 1987-2000. *Journal of the American Medical Association 292*(19), 2372.

Berger, J., B. Liseo, and R. Wolpert (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*.

Bobb, J. F., F. Dominici, and R. D. Peng (2011). A Bayesian model averaging approach for estimating the relative risk of mortality associated with heat waves in 105 U.S. cities. *Biometrics*.

Chen, J., D. Zhang, and M. Davidian (2002). A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics 3*, 347–360.

Dominici, F., R. D. Peng, M. L. Bell, L. Pham, A. McDermott, S. L. Zeger, and J. M. Samet (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association 295*(10), 1127–34.

COBRA Collection of Biostatistics Research Archive

Dominici, F., J. M. Samet, and S. L. Zeger (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society, Series A 163*(3), 263–302.

Edwards, A. (1992). *Likelihood*. Baltimore: Johns Hopkins University Press.

Efron, B. (1996). Empirical Bayes methods for combining likelihoods. *Journal of the American Statistical Association 91*(434), 538–565.

EPA (2011). The Benefits and Costs of the Clean Air Act from 1990 to 2020. Technical report, U.S. Environmental Protection Agency Office of Air and Radiation, Washington, D.C.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics 1*(2), 209–230.

Gallant, A. R. and D. W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica 55*(2), 363–390.

Heagerty, P. and B. Kurland (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika 88*(4), 973.

Katsouyanni, K., G. Touloumi, E. Samoli, A. Gryparis, A. L. Tertre, Y. Monopolis, G. Rossi, D. Zmirou, F. Ballester, and A. Boumghar (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 european cities within the aphea2 project. *Epidemiology 12*(5), 521.

Komárek, A. and E. Lesaffre (2008). Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Computational Statistics & Data Analysis 52*(7), 3441–3458.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association 73*(364), 805–811.

Liao, J. (1999, Jan). A hierarchical Bayesian model for combining multiple 2 x 2 tables using conditional likelihoods. *Biometrics 55*(1), 268–272.

Litière, S., A. Alonso, and G. Molenberghs (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine 27*, 3125–3144.

Litière, S., A. Alonso, and G. Molenberghs (2010). Rejoinder to "A note on type II error under random effects misspecification in generalized linear mixed models". *Biometrics*.

Magder, L. S. and S. L. Zeger (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *Journal of the American Statistical Association 91*(435), 1141–1151.

McCulloch, C. E. and J. M. Neuhaus (2011). Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics 67*(1), 270–279.

BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics 9*(2), 249–265.

Neuhaus, J., W. Hauck, and J. Kalbfleisch (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika 79*(4), 755–762.

NRC (2004). Research priorities for airborne particulate matter: IV. Continuing research progress. Technical report, National Research Council of the National Academies, Washington, D.C.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press.

Peng, R., H. Chang, M. L. Bell, A. McDermott, S. L. Zeger, J. M. Samet, and F. Dominici (2008). Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients. *Journal of the American Medical Association 299*(18), 2172–2179.

Peng, R. D., M. L. Bell, A. S. Geyh, A. McDermott, S. L. Zeger, J. M. Samet, and F. Dominici (2009). Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. *Environmental Health Perspectives 117*(6), 957–963.

Schwartz, J. and B. A. Coull (2003). Control for confounding in the presence of measurement error in hierarchical models. *Biostatistics 4*(4), 539–553.

Verbeke, G. and E. Lesaffre (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis 23*, 541–556.

Zeka, A. and J. Schwartz (2004). Estimating the independent effects of multiple pollutants in the presence of measurement error: an application of a measurement-error–resistant technique. *Environmental Health Perspectives 112*(17), 1686–1690.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Table 1: Summarizes simulation results for the cluster-specific log odds ratios θ_i, showing the squared error loss $\sum_{i=1}^{I}(\tilde{\theta}_i - \theta_i)^2$, along with coverage of 95% posterior intervals; Displays results for the mean log odds ratio θ^*, showing bias, standard deviation, rMSE, and coverage of 95% posterior intervals. Methods compared are the full Bayesian hierarchical model (FBHM), reduced BHM with conditional likelihood and normal random-effect distribution (RBHM-L-N), reduced BHM with normal approximation to the likelihood and normal random-effect distribution (RBHM-N-N), and reduced BHM with conditional likelihood and Dirichlet-Process normal mixture for the random-effect distribution (RBHM-L-DP).

	Simulation	**Cluster** θ_i		**Overall** θ^*			
		Sq. Error	Coverage	Bias	SD	rMSE	Coverage
	1(a)*				$\theta^* = -1.3$		
	(i) FBHM	14.4	0.95	0.00	0.11	0.11	0.94
	(ii) RBHM-L-N	14.8	0.95	0.02	0.11	0.11	0.95
	(iii) RBHM-L-DP	14.8	0.95	0.03	0.11	0.11	0.94
Model 1:	(iv) RBHM-N-N	18.0	0.94	0.09	0.10	0.14	0.89
Bivariate	**1(b)***				$\theta^* = -1.3$		
Normal	(i) FBHM	14.9	0.95	−0.01	0.12	0.12	0.94
	(ii) RBHM-L-N	18.9	0.94	0.04	0.11	0.12	0.93
	(iii) RBHM-L-DP	18.9	0.94	0.06	0.11	0.12	0.92
	(iv) RBHM-N-N	27.5	0.92	0.14	0.10	0.17	0.74
	2(a)				$\theta^* = 1.46$		
	(i) FBHM	7.5	0.88	−0.03	0.04	0.05	0.87
	(ii) RBHM-L-N	8.0	0.91	−0.03	0.04	0.05	0.90
	(iii) RBHM-L-DP	7.5	0.95	−0.04	0.04	0.06	0.90
Model 2:	(iv) RBHM-N-N	9.6	0.89	−0.07	0.04	0.08	0.66
Uniform/	**2(b)**				$\theta^* = 1.67$		
Beta	(i) FBHM	99.5	0.90	−0.14	0.08	0.16	0.55
	(ii) RBHM-L-N	104.2	0.91	−0.13	0.08	0.16	0.56
	(iii) RBHM-L-DP	96.2	0.92	−0.14	0.09	0.17	0.57
	(iv) RBHM-N-N	137.6	0.89	−0.23	0.07	0.24	0.11
	3(a)				$\theta^* = -1.3$		
	(i) FBHM	10.6	0.95	−0.01	0.11	0.11	0.97
	(ii) RBHM-L-N	11.6	0.95	0.00	0.11	0.11	0.97
	(iii) RBHM-L-DP	9.8	0.96	0.02	0.10	0.11	1.00
Model 3:	(iv) RBHM-N-N	11.6	0.95	0.06	0.10	0.12	0.94
Normal	**3(b)**				$\theta^* = -1.3$		
Mixture	(i) FBHM	10.1	0.95	−0.02	0.12	0.13	0.93
	(ii) RBHM-L-N	13.5	0.95	0.02	0.12	0.12	0.94
	(iii) RBHM-L-DP	12.1	0.96	0.03	0.12	0.12	0.99
	(iv) RBHM-N-N	15.0	0.94	0.12	0.11	0.17	0.80
	4(a)				$\theta^* = 0.8$		
	(i) FBHM	7.9	0.84	0.00	0.06	0.06	0.85
	(ii) RBHM-L-N	7.2	0.93	0.01	0.06	0.06	0.94
	(iii) RBHM-L-DP	7.3	0.97	−0.01	0.06	0.06	0.96
Model 4:	(iv) RBHM-N-N	7.5	0.90	−0.05	0.05	0.07	0.86
Normal-θ_i	**4(b)**				$\theta^* = 0.8$		
	(i) FBHM	5.1	0.93	0.00	0.05	0.05	0.93
	(ii) RBHM-L-N	5.2	0.94	0.00	0.05	0.05	0.94
	(iii) RBHM-L-DP	5.2	0.96	−0.01	0.04	0.05	0.94
	(iv) RBHM-N-N	5.2	0.94	−0.02	0.04	0.05	0.93

*For scenarios 1(a) and 1(b), the summary statistics for RBHM-L-DP are based on 999 and 998 simulation repetitions, respectively. The other repetitions were excluded because the MCMC didn't converge within the maximum number of iterations.

Hosted by The Berkeley Electronic Press