Prenatal Exposure to Correlated Environmental Contaminants

Franco Momoli & Jillian Ashley Martin
SPER Advanced Methods Workshop
June 15, 2015

Workshop Objectives

• To provide an overview of Bayesian hierarchical models and how they can be used to account for correlated exposures
• To introduce software programs for running Bayesian hierarchical models
• To illustrate use of Bayesian hierarchical models using an example from a Canadian birth cohort study
• To articulate the advantages and limitations of these models

Overview

Hierarchical Regression:

• Parameter estimates treated as random variables and described according to prior distribution
• Shrinkage estimates: values shrunk away from maximum likelihood estimate and towards mean of prior distribution

\[ \beta_j \sim \mathcal{N}(\mu, \Phi^2) \]

- \( \mu \): prior knowledge about true value of \( \beta \)
- \( \Phi^2 \): uncertainty regarding \( \mu \) (prior variance)

Degree of shrinkage depends on \( \Phi^2 \)

Large variance \( \Phi^2 \): greater uncertainty about prior \( \mu \) less shrinkage

How to account for multiple exposures?

• Bayesian approach
• Application to Canadian birth cohort study
  - Example: Prenatal environmental contaminant exposure \( \rightarrow \) cord blood levels of IgE

Shrinkage

\( \Theta \): Target parameter

Rifle 1: Unbiased, Large scatter (random error)
Rifle 2: Moderate bias, moderate random error
Rifle 3: Large bias, Low random error

How can we improve our estimate of \( \Theta \)?

Figure 1: Observers of shots (trajectories) from three different rifles (regardless on either) in \( m \) = Rifle 1 shots, \( n \) = Rifle 2 shots, \( k \) = Rifle 3 shots.
**Shrinkage**

- Best guess about value of $\theta$ (prior mean)

Each bullet from Rifle 1 is deflected half-way towards $r$.

Slight increase in bias, decrease in scatter:
- On average, our estimate is closer to $\theta$.
- Final estimate of target parameter combines prior mean with observed data.

**Bayesian Hierarchical Models**

Differ according to how prior distribution is specified

- **Semi-Bayes**
  - Example prior for $B_i \sim N(\mu, \phi^2)$
  - Fixed variance ($\phi^2$ Constant)

- **Bayes**
  - Example prior for $B_i \sim N(\mu, \phi^2)$
  - $\phi^2$ not fixed
  - Distributions
    - Inverse gamma
    - Half-normal
    - Uniform

- Semi-parametric Bayes
  - Not subject to assumptions inherent in use of normally distributed prior
  - Clusters coefficients on effect size rather than based on researcher choice

---

**The Maternal Infant Research on Environmental Chemicals (MIREC) Study**

- Recruitment from 10 sites across the country between 2008-2011
- Inclusion:
  - Age 18 years or over
  - <14 weeks gestation
  - Willing to provide a cord blood sample
  - Planning on delivering at local hospital

**Objectives**

- Prenatal exposure to:
  - Bisphenol A (BPA)
  - Phthalates
  - Perfluoralkyl substances

**Rationale**

- Elevated levels of IgE are a biomarker of childhood allergy
- Environmental contaminants hypothesized to contribute to risk of childhood allergy

**MIREC Study Design**

- Pregnancy Trimester
  - Recruitment
  - $N=2001$
  - Cord Blood
  - $N=1363$

**Exposures**

- Phthalates
- Bisphenol A
- Perfluoralkyl substances
- Metals
- Persistent organic pollutants

**Questionnaire/Data**

- Reproductive and demographic characteristics
Environmental Chemicals & Sources of Exposure

<table>
<thead>
<tr>
<th>Chemical Class</th>
<th>Metabolites/Chemicals</th>
<th>Sources of Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phthalates</td>
<td>MEP</td>
<td>PVC products (flooring)</td>
</tr>
<tr>
<td></td>
<td>MBP</td>
<td>Medical tubing</td>
</tr>
<tr>
<td></td>
<td>MCPP</td>
<td>Household (consumer products)</td>
</tr>
<tr>
<td></td>
<td>MEHP</td>
<td>Nail polish, cosmetics, fragrances</td>
</tr>
<tr>
<td></td>
<td>MEOHP</td>
<td>Plastics</td>
</tr>
<tr>
<td></td>
<td>MEHHP</td>
<td>Plastics</td>
</tr>
<tr>
<td></td>
<td>MBzP</td>
<td>Plastic</td>
</tr>
<tr>
<td></td>
<td>BPA</td>
<td>Plastics</td>
</tr>
<tr>
<td></td>
<td>MEHP</td>
<td>Canned goods (epoxy resin)</td>
</tr>
<tr>
<td></td>
<td>MEOHP</td>
<td>Receipts (thermal printing paper)</td>
</tr>
</tbody>
</table>

Correlated Exposures

Pearson correlation coefficient of log_{10} transformed phthalate metabolites and BPA

<table>
<thead>
<tr>
<th>MEP</th>
<th>MBP</th>
<th>MCPP</th>
<th>MEHP</th>
<th>MEOHP</th>
<th>MEHHP</th>
<th>BPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEP</td>
<td>1.03</td>
<td>0.37</td>
<td>0.35</td>
<td>0.40</td>
<td>0.40</td>
<td>0.10</td>
</tr>
<tr>
<td>MBP</td>
<td>1.70</td>
<td>1.63</td>
<td>0.60</td>
<td>0.66</td>
<td>0.46</td>
<td>0.52</td>
</tr>
<tr>
<td>MCPP</td>
<td>0.58</td>
<td>0.65</td>
<td>0.49</td>
<td>0.70</td>
<td>0.52</td>
<td>0.40</td>
</tr>
<tr>
<td>MEHP</td>
<td>0.52</td>
<td>0.31</td>
<td>0.45</td>
<td>0.45</td>
<td>0.31</td>
<td>0.40</td>
</tr>
<tr>
<td>MEOHP</td>
<td>1.09</td>
<td>0.49</td>
<td>0.47</td>
<td>0.47</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>MEHHP</td>
<td>0.49</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>BPA</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>PFOS</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

All within chemical class correlations p-value <0.01

Factors to consider in modeling strategy

• Can reasonable clusters of exposures (priors) be identified?

Correlated Exposures

Pearson correlation coefficient of log_{10} transformed perfluoroalkyl substances

<table>
<thead>
<tr>
<th>PFOS</th>
<th>PFOA</th>
<th>PFHxS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFOS</td>
<td>1.00</td>
<td>0.58</td>
</tr>
<tr>
<td>PFOA</td>
<td>0.54</td>
<td>0.47</td>
</tr>
<tr>
<td>PFHxS</td>
<td>0.54</td>
<td>0.47</td>
</tr>
</tbody>
</table>

All correlations significant at p<0.01, n=1248

Factors to consider when identifying priors

• Statistical correlation
  • Within chemical class correlation greater than between chemical class correlation
  • MEHP, MEHHP, MEOHP = Metabolites of DEHP (same parent compound) - ∑ DEHP

• Toxicology & chemical properties
• Potential immunotoxic mechanisms
• Exchangeability
  • Coefficients for different chemical classes originate from different prior distributions
Models

• #1: Phthalate metabolites & BPA → IgE
  • #2 Perfluoroalkyl substances → IgE

Phthalate & BPA hierarchical logistic regression model

\[ Y_i | \beta_j \sim N(\sum \beta_j x_{ij}, \sigma^2) \]

Underlying logistic regression model was of the form:

\[ \text{Logit}(\text{Pr}(\text{IgE}=1|\text{MEP} + \text{MBP} \ldots)) = B_0 + B_1 \cdot \text{MEP} + B_2 \cdot \text{MBP} + B_3 \cdot \text{MBzP} + B_4 \cdot \sum \text{DEHP} + B_5 \cdot \text{BPA} + B_6 \cdot \text{MCPP} + B_7 \cdot \text{MCPP} \cdot \text{MCPP} + \text{specific gravity} + \text{age} \]

IgE = 1 represents IgE > 0.5 \( \mu \text{L}^{-1} \)

\( \sum \text{DEHP} = \text{summary index of MEHP, MEHHP, MEOHP} \)

MCPP*MCPP included as spline analysis showed non-linear relationship between MCPP and IgE

Specific gravity included as covariate to account for heterogeneity in urinary dilution

Prior Distributions:

\[ \beta \sim \text{N}(0, \Phi \sigma^2) \]

\( \Phi \sigma^2 \sim \text{Half-N}(0,0.01) \)

\[ \text{Age} \sim \text{N}(0,0.01) \]

Specific gravity included as covariate to account for heterogeneity in urinary dilution

If prior information available for confounder-outcome relationship, this can be reflected in prior distribution (e.g. Smoking as a confounder in a model with birth weight as the outcome).

Software

• R Studio and Openbugs
  • R Studio and Openbugs
  • R Studio and Openbugs
  • Other programs
    • JAGS  (just another gibbs sampler)
    • R package
    • Rjags package
    • http://mcmc-jags.sourceforge.net/

Openbugs output

<table>
<thead>
<tr>
<th>Parameter</th>
<th>mean</th>
<th>sd</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>n.eff</th>
<th>Rhat</th>
<th>ESS.eff</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.918</td>
<td>0.107</td>
<td>0.7247</td>
<td>0.9135</td>
<td>1.142</td>
<td>9000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORx1</td>
<td>0.907</td>
<td>0.181</td>
<td>0.697</td>
<td>0.9899</td>
<td>1.408</td>
<td>9000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORx2</td>
<td>0.9905</td>
<td>0.160</td>
<td>0.7093</td>
<td>0.9789</td>
<td>1.34</td>
<td>9000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORx3</td>
<td>0.9237</td>
<td>0.141</td>
<td>0.6813</td>
<td>0.913</td>
<td>1.23</td>
<td>9000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORx4</td>
<td>0.8984</td>
<td>0.153</td>
<td>0.6244</td>
<td>0.8877</td>
<td>1.227</td>
<td>9000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORx5</td>
<td>1.002</td>
<td>0.149</td>
<td>0.7357</td>
<td>0.9915</td>
<td>1.321</td>
<td>9000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
R Output

- Inference for Bugs model at "paa_adj_linear_5_5.txt",
- Current: 3 chains, each with 10000 iterations (first 500 discarded)
- Cumulative: n.sims = 28500 iterations saved

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
<th>Rhat</th>
<th>n.eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>-0.2</td>
<td>0.1</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>1</td>
<td>6800</td>
</tr>
<tr>
<td>[2]</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>1</td>
<td>1500</td>
</tr>
<tr>
<td>[3]</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>1</td>
<td>3900</td>
</tr>
</tbody>
</table>

Rhat = 1 = convergence (when chains have 'forgotten' their initial values and output is indistinguishable); n.eff = crude measure of effective sample size

Model Comparison

<table>
<thead>
<tr>
<th>Log_{10} phthalate metabolites (ug/L)</th>
<th>Logistic Regression Models CR (95% CI) High IgE</th>
<th>Bayesian Hierarchical Model CR (95% CI) High IgE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEP</td>
<td>0.8 [0.6 - 1.0]</td>
<td>0.8 [0.6 - 1.0]</td>
</tr>
<tr>
<td>MBF</td>
<td>1.0 [0.7 - 1.5]</td>
<td>0.9 [0.6 - 1.3]</td>
</tr>
<tr>
<td>MBzP</td>
<td>1.3 [0.9 - 1.8]</td>
<td>1.2 [0.9 - 1.7]</td>
</tr>
<tr>
<td>MCPP</td>
<td>0.8 [0.6 - 1.2]</td>
<td>1.2 [0.9 - 1.5]</td>
</tr>
<tr>
<td>Σ DEHP</td>
<td>1.0 [0.8 - 1.4]</td>
<td>1.0 [0.7 - 1.5]</td>
</tr>
<tr>
<td>BPA</td>
<td>1.0 [0.7 - 1.5]</td>
<td>1.0 [0.7 - 1.3]</td>
</tr>
</tbody>
</table>

*all models adjusted for age and specific gravity
n=1136

Convergence...

Conclusions

- No observed association between either phthalates or perfluoroalkyl substances and IgE in either model
- Bayesian results were comparable to single chemical models
  - Some ORs closer to null (shrinkage)
  - Slightly tighter credible intervals
- Bayesian hierarchical model was a feasible approach for accounting for correlated exposures

Work in progress:
Perfluoroalkyl substances and birth weight; MIREC study

PFOA  PFS  Birth weight

Rationale: previous literature & meta-analysis
Work in progress: Perfluoroalkyl substances and birth weight: MIREC study

• Rationale
  • Recent meta-analysis reported 18 g decrease in birth weight per 1 ng/mL increase in maternal PFOS levels.

  • Incorporate information into prior distributions
  • Influence of correlated exposures

Bayesian Hierarchical Models

• Advantages
  • Ability to control for correlated exposures without model instability or lack of convergence
  • Minimizes type 1 error by shrinking parameter estimates to prior mean
  • Avoids 'single chemical' approaches
  • Mean squared error may be lower than maximum likelihood estimates

• Limitations
  • Do not account for cumulative exposures or possible synergism between chemicals
  • Assumes 'exchangeability' within clusters
  • Parametric models reliant on researcher defined clusters (prions)
  • Initial effort more intense

Tips

• Test different prior distributions
• Use multiple approaches to identify data clusters
  • Biological rationale
    • Chemical class
    • Toxicological activity (e.g. endocrine disruptor)
    • Statistical correlation

  • Test multiple priors

References

• Books
  • Bayesian Data Analysis by Andrew Gelman
  • Bayesian Approaches to Clinical Trials and Health Care Evaluations by Dr. Spiegelhalter

• Articles
  • MacDonell et al. Bayesian methods for highly correlated exposure data. Epidemiology 2007; 18:191-197

• Websites
  • Lawrence Joseph at McGill teaching websites: http://www.medicine.mcgill.ca/epidemiology/epi668/CourseOutline.html
  • Openbugs tutorial: http://mcdougall.sfu.ca/teaching/BI204/LectureNotes/lecture11.pdf
  • Openbugs Manual: http://mcdougall.sfu.ca/teaching/BI204/LectureNotes/TutorialOpenbugs.html

Acknowledgements

Mentors: Dr. Linda Dodds
Dr. Rob Platt
Dr. Adrian Levy
Dr. Tye Arbuckle
Dr. Jean Marshall

Support: Members of Marshall Lab
Nicole Lupien, MIREC Biobank Committee
MIREC study participants

Thank you!

jillian.ashley-martin@dal.ca

Example Openbugs Code

```r
model {
  for (i in 1:n) {
    # Linear regression on logit
    logit(y[i]) <- alpha + b[1]*x1[i]  + b[2]*x2[i] + b[3]*x3[i]
    # Likelihood function for each data point
    y[i] ~ dbern(p[i])
  }
  for (j in 1:J) {
    b[j] ~ dnorm(0,phi) # Prior for betas
  }
  alpha ~ dnorm(0.0,0.01) # Prior for intercept
  phi  ~ dgamma(0.3,1) # Inverse gamma or half-normal (dnorm(0,0.01)I(0,))
  # Calculate odds ratios for three betas
  ORx1<-exp(b[1])
  ORx2<-exp(b[2])
  ORx3<-exp(b[3])
}
```