Prenatal Exposure to Correlated Environmental Contaminants

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

1

How to account for multiple exposures?

- Bayesian approach
- Application to Canadian birth cohort study
  - Example: Prenatal environmental contaminant exposure → cord blood levels of IgE

2

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

3

Bayesian Methods for Highly Correlated Exposure Data

Richard F. MacLehose,† David B. Dunson,† Amy H. Herring,‡§ and Jane A. Hoppin¶*

Epidemiology • Volume 18, Number 2, March 2007

4

Overview

Hierarchical Regression:

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

$$\beta j \sim N(\mu, \Phi^2)$$

μ = prior knowledge about true value of β
Φ^2 = uncertainty regarding μ (prior variance)

Degree of shrinkage depends on Φ^2
Large variance → greater uncertainty about prior → less shrinkage

5

Shrinkage

Θ = 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 Θ?

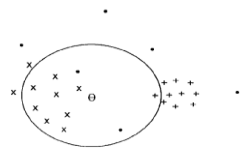


Figure 1 Clusters of shots (estimates) from three different rifles (estimators) sighted on point **Θ**: • = Rifle 1 shots, X = Rifle 2 shots, + = Rifle 3 shots

Principles of multilevel modelling
Sander Greenland

International Journal of Epidemiology 2000;29:158–167

6

Shrinkage

r = best guess about value of θ (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 θ.
Final estimate of target parameter combines prior mean with observed data.

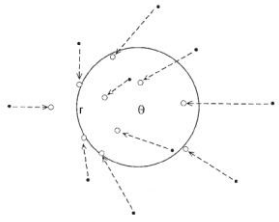


Figure 2 How cluster from Rifle 1 could be made better by pulling toward a point r that need not equal θ

Principles of multilevel modelling

International Journal of Epidemiology 2000;29:158–167

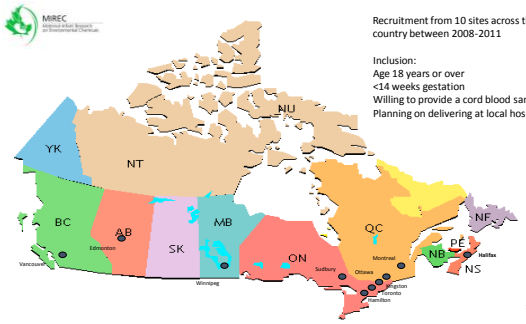
Bayesian Hierarchical Models

Differ according to how prior distribution is specified

- Semi-Bayes
  - Example prior for Bj ~ N(μ, Φ^2)
  - Fixed variance (Φ^2 ~ Constant)
- Bayes
  - Example prior for Bj ~ N(μ, Φ^2)
  - Φ^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

Maclehose et al 2007

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
-Perfluoroalkyl substances

→ Elevated umbilical cord blood levels of Immunoglobulin E (IgE)

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	1st →	2nd →	3rd →	Cord Blood N=1363 ⇢
Exposures	**Phthalates** **Bisphenol A** **Perfluoroalkyl substances** Metals Persistent organic pollutants			
Questionnaire Data	Reproductive and demographic characteristics			

Environmental Chemicals & Sources of Exposure

Chemical Class	Metabolites/Chemicals		Sources of Exposure
Phthalates	MEP MBzP MEHP MEOHP	MBP MCPP MEHHP	•PVC products (flooring) •Medical tubing •Household /consumer products (Nail polish, cosmetics, fragrances) •Plastics
Bisphenol A			•Plastics •Canned goods (epoxy resin) •Receipts (thermal printing paper)

13

Factors to consider in modeling strategy

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

14

Correlated Exposures

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

	MEP	MBP	MBzP	MCPP	MEHP	MEOHP	MEHHP	BPA	PFOA	PFOS	PFHxS
MEP	1	0.43	0.39	0.37	0.35	0.40	0.40	0.30	0.05	0.02	-0.03
MBP		1	0.70	0.63	0.60	0.66	0.66	0.52	-0.08	-0.01	-0.04
MBzP			1	0.57	0.54	0.59	0.60	0.46	-0.04	0.03	0.02
MCPP				1	0.58	0.65	0.67	0.49	-0.08	0.00	-0.03
MEHP					1	0.92	0.91	0.45	-0.12	0.01	-0.06
MEOHP						1	0.99	0.49	-0.07	0.01	-0.07
MEHHP							1	0.49	-0.07	0.01	-0.07
BPA								1	-0.04	0.03	0.02
PFOA									1	0.55	0.46
PFOS										1	0.52
PFHxS											1

All within chemical class correlations p-value <0.01

15

Correlated Exposures

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

	MEP	MBP	MBzP	MCPP	MEHP	MEOHP	MEHHP	BPA
MEP	1	0.43	0.39	0.37	0.35	0.40	0.40	0.30
MBP		1	0.70	0.63	0.60	0.66	0.66	0.52
MBzP			1	0.57	0.54	0.59	0.60	0.46
MCPP				1	0.58	0.65	0.67	0.49
MEHP					1	0.92	0.91	0.45
MEOHP						1	0.99	0.49
MEHHP							1	0.49
BPA								1

All correlations significant at p<0.01, n=1151

16

Correlated Exposures

Pearson correlation coefficient of $\log_e$ transformed perfluoroalkyl substances

	PFOA	PFOS	PFHxS
PFOA	1	0.58	0.47
PFOS		1	0.54
PFHxS			1

All correlations significant at p<0.01, n=1248

17

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

18

Models

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

Phthalate & BPA hierarchical logistic regression model

$$Y_i|\beta_j \sim N(\sum_{J=1}^{k} \beta x_j, \sigma^2)$$

$\beta \sim N(0, \Phi^2)$
$\Phi^2 \sim$ Half- N (0,0.01)

Underlying logistic regression model was of the form:

Logit Pr(IgE=1|MEP + MBP …) = Bo + B1*MEP + B2*MBP + B3*MBzP + B4*∑DEHP + B5*BPA + B6*MCPP + B7*MCPP*MCPP + specific gravity + age

IgE = 1 represents IgE > 0.5 ku/L
- ∑DEHP = 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:

$\beta \sim N(0, \Phi^2)$ = prior distribution for parameter estimates
Mean = 0 due to lack of prior literature regarding exposures and cord blood IgE

$\Phi^2 \sim$ Half-N (0,0.01) = prior distribution of variance of parameter estimates
Note: Openbugs uses precision rather than variance
Φ^2 is uninformative due to large variance

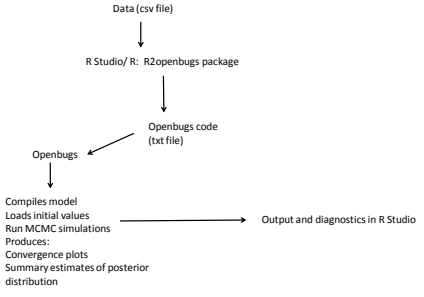
Specific gravity ~N(0,0.01)
Age ~N(0,0.01) = prior distribution for covariates
Uninformative due to lack of information regarding covariate-IgE relationship

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 package – R2OpenBugs
  - Allows interface between R and Openbugs
- Other programs
  - JAGS (just another gibbs sampler)
    - Rjags package
    - http://mcmc-jags.sourceforge.net/
  - STAN
    - Rstan package
    - http://mc-stan.org/tutorials.html

Process



Openbugs output

model is syntactically correct
data loaded
model compiled
initial values generated, model initialized
model is updating
500 updates took 11 s
model is updating
4500 updates took 80 s
CODA files written
Summary statistics

	mean	sd	val2.5pc	median	val97.5pc	sample
ORx1	0.918	0.1071	0.7247	0.9135	1.142	9000
ORx2	1.007	0.1813	0.697	0.9899	1.408	9000
ORx3	0.9905	0.1604	0.7093	0.9789	1.34	9000
ORx4	0.9237	0.1416	0.6813	0.913	1.23	9000
ORx5	0.8984	0.1538	0.6244	0.8877	1.227	9000
ORx6	1.002	0.1499	0.7357	0.9915	1.321	9000
alpha	-1.366	0.2869	-1.899	-1.37	-0.8015	9000
b[1]	-0.09233	0.1165	-0.3211	-0.09039	0.1331	9000
b[2]	-0.008811	0.1776	-0.3611	-0.009856	0.3432	9000
b[3]	-0.02256	0.1615	-0.3428	-0.02134	0.2936	9000
b[4]	-0.091	0.1525	-0.3843	-0.09101	0.2072	9000
b[5]	-0.1218	0.1717	-0.4716	-0.1191	0.2073	9000
b[6]	-0.009423	0.1493	-0.3058	-0.008437	0.2783	9000
deviance	1005.0	2.83	1001.0	1005.0	1012.0	9000

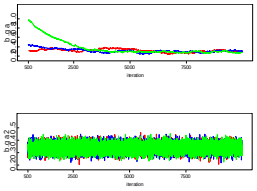
R Output

- 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

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
b[1]	-0.2	0.1	-0.4	-0.2	-0.2	-0.1	0.1	1	6800
b[2]	0.1	0.1	-0.1	0	0.1	0.2	0.3	1	1500
b[3]	0	0.1	-0.1	0	0	0.1	0.1	1	3900

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

Convergence …?



Model Comparison

Log10 phthalate metabolites (ug/L)	Logistic Regression Models OR (95% CI) High IgE	Bayesian Hierarchical Model OR (95% CI) High IgE
MEP	0.8 (0.6-1.0)	0.8 (0.6-1.0)
MBP	1.0 (0.7-1.5)	0.9 (0.6-1.3)
MBZP	1.3 (0.9-1.8)	1.2 (0.9-1.7)
MCPP	0.8 (0.6- 1.2)	0.7 (0.5-1.0)
∑ DEHP	1.1 (0.7-1.6)	1.0 (0.7-1.5)
BPA	1.0 (0.7-1.5)	1.0 (0.7-1.3)

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

Ashley-Martin et al, Environ Res. 2015; 140:360-368.

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
PFOS
PFHxS

→ Birth weight

Rationale: previous literature & meta-analysis

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Integration of Animal and Human Evidence for PFOA Effects on Fetal Growth

Juleen Lam,[1] Erica Koustas,[1] Patrice Sutton,[2] Paula I. Johnson,[2] Dylan S. Atchley,[2] Saunak Sen,[3] Karen A. Robinson,[4,5,6] Daniel A. Axelrad,[7] and Tracey J. Woodruff[2]

VOLUME 122 | NUMBER 10 | October 2014 • Environmental Health Perspectives

RESULTS: We identified 18 human studies that met our inclusion criteria, and 9 of these were combined through meta-analysis. Through meta-analysis, we estimated that a 1-ng/mL increase in serum or plasma PFOA was associated with a –18.9 g (95% CI: –29.8, –7.9) difference in birth weight. We concluded that the risk of bias across studies was low, and we assigned a "moderate" quality rating to the overall body of human evidence.

CONCLUSION: On the basis of this first application of the Navigation Guide systematic review methodology, we concluded that there is "sufficient" human evidence that developmental exposure to PFOA reduces fetal growth.

CONCLUSION: We concluded that developmental exposure to PFOA adversely affects human health based on sufficient evidence of decreased fetal growth in both human and nonhuman mammalian species. The results of this case study demonstrate the application of a systematic and transparent methodology, via the Navigation Guide, for reaching strength of evidence conclusions in environmental health.

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 PFOA 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 (priors)
  - 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 DJ Spiegelhalter
- Articles
  - Maclehose et al. Bayesian methods for highly correlated exposure data. Epidemiology 2007; 18:199-207
- Websites
  - Lawrence Joseph at McGill teaching websites
    http://www.medicine.mcgill.ca/epidemiology/Joseph/courses/EPIB-668/CourseOutline.html
  - R Tutorial. An Introduction to Statistics
    http://www.r-tutor.com/bayesian-statistics/openbugs
  - R2Winbugs package description file
    http://cran.r-project.org/web/packages/R2WinBUGS/R2WinBUGS.pdf
  - Openbugs Manual
    http://mathstat.helsinki.fi/openbugs/Manuals/Tutorial.html

Acknowledgements

Mentors: Dr. Linda Dodds, Dr. Adrian Levy, Dr. Rob Platt, 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

IWK Health Centre, Health Research Foundation, CIHR IRSC, Health Canada, Santé Canada

Example Openbugs Code

```
model
{
for (i in 1:n) {

#Linear regression on logit
logit(p[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
Or1<-exp(b[1])
Or2<-exp(b[2])
Or3<-exp(b[3])
}
```