

# INTRODUCTION TO BAYESIAN HIERARCHICAL METHODS

## MULTIPLE EXPOSURE MODELING

ALPINE EXPOSOME SUMMER SCHOOL  
JUNE 2021

Patrick T. Bradshaw, Ph.D.  
Associate Professor  
Division of Epidemiology  
University of California, Berkeley

Berkeley Public  
Health

# OUTLINE

Challenge of Multiple Exposures

- Hierarchical modeling

- Insight from Mixed Models

- Interpretation of estimates

Application: Pesticides and Non-Hodgkin's Lymphoma

Final thoughts

References

# CHALLENGE OF MULTIPLE EXPOSURES

# THE CHALLENGE OF MULTIPLE EXPOSURES

Complex exposures are common in many areas of epidemiology:

- **Environmental health:** simultaneous exposure to multiple chemicals; possibly overlapping effects.
- **Nutrition:** diet consists of multitude of foods consumed, each with a host of nutrients/constituents.
- **Genetics:** we possess many genes, operating on related biological pathways.

*How to model this complex array of exposures?*



Photo by Girl with red hat on Unsplash

# THE CHALLENGE OF MULTIPLE EXPOSURES

Multiple exposure modeling is an active area of research.

- Emphasis area for the U.S. National Institute of Environmental Health Sciences (<https://www.niehs.nih.gov/research/supported/exposure/mixtures/index.cfm>).
- Symposium at The 2018 Annual Meeting of the *Society for Epidemiologic Research* to discuss approaches in nutrition and environmental epidemiology.
- Growing number of methodological publications in this area. Nice review:

Current Epidemiology Reports  
<https://doi.org/10.1007/s40471-018-0145-0>

EPIDEMIOLOGIC METHODS (R MACLEHOSE, SECTION EDITOR)



## Environmental Exposure Mixtures: Questions and Methods to Address Them

Ghassan B. Hamra<sup>1</sup> • Jessie P. Buckley<sup>1,2</sup>

# THE CHALLENGE OF MULTIPLE EXPOSURES

**Serial analyses** (multiple models, single exposure at-a-time):

- Ignores other relevant factors:
  - Confounding by other related exposures/residual effects.
- Search for “magic pill” is major limitation in epidemiology.
- Contributes to inconsistency in epidemiologic literature.



*The American Journal of Clinical Nutrition*

Is everything we eat associated with cancer? A systematic  
cookbook review<sup>1–3</sup>

Jonathan D Schoenfeld and John PA Ioannidis

**Conclusions:** Associations with cancer risk or benefits have been claimed for most food ingredients. Many single studies highlight implausibly large effects, even though evidence is weak. Effect sizes shrink in meta-analyses. *Am J Clin Nutr* 2013;97:127–34.

# THE CHALLENGE OF MULTIPLE EXPOSURES

**Simultaneous multiple exposure modeling** → sparse data.

- Wide confidence intervals.
- Non-estimable parameters.

**Empirically-driven data reduction methods:**

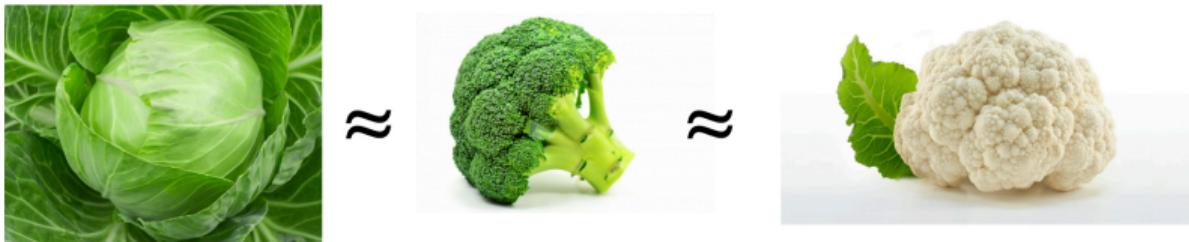
- Factor analysis/Principal components analysis.
- Cluster analysis/Latent class analysis.
- Reduced rank regression.

*Interpretability?*

# LEVERAGING WHAT YOU KNOW

We often have expectations (priors) for how **exposures** operate:

- Similar *nutrient compositions* → similar effects on disease risk.
- Similar **chemical structure** → similar effects on disease risk.
- Similar **metabolic pathways** → similar effects on disease risk.



# HIERARCHICAL MODELING

- Sensible to “shrink” effects of similar exposures closer together.
  - **Grouping like exposures:** motivation for aggregate (summed) exposures, dietary pattern scores, etc....
- **Hierarchical modeling** can formalize this.
  - Introduced to epidemiologic applications by Greenland (*Stat Med* 1993).
  - Subsequent development/variations (see references).
- **Numerous applications:**
  - Environmental epidemiology.
  - Nutritional epidemiology.
  - Genetic epidemiology.

# HIERARCHICAL MODELING

## A Bayesian hierarchical model:

- Broad term referring to case when there is a hierarchical (multilevel) structure between model parameters.
- Often useful when need to borrow information across variables when estimating many parameters.
- Wide range of applications.

# HIERARCHICAL MODEL SPECIFICATION

## Bayesian model specification (GLM framework):

1. **Outcome model** (likelihood):  $p(y)$ ; note set of exposures  $\mathbf{x} = \{x_j\}_{j=1}^J$  and additional covariates (e.g. confounders)  $\mathbf{w} = \{w_l\}_{l=1}^L$ :  
For example, a logistic regression model:

$$\text{logit}(E[Y|\mathbf{x}, \mathbf{w}]) = \alpha + \sum_j x_j \beta_j + \sum_l w_l \gamma_l$$

(in practice may be applied to any regression model with linear link function).

# HIERARCHICAL MODEL SPECIFICATION

2. **Model for exposure effects:** often (but not necessarily) Normal (Gaussian), mean is dependent on variables (e.g. characteristics of exposures given by  $\mathbf{z}_j = \{z_{jk}\}_{k=1}^K$ ):

$$\begin{aligned}\beta_j &= \sum_k z_{jk} \pi_k + \delta_j \\ \delta_j &\sim N(0, \sigma^2)\end{aligned}$$

- If exposures are all equally similar (no unique characteristics), can let  $\mathbf{z} = 1$  (intercept only, or shared mean model).

**Hierarchical** because distribution of  $\beta$  depends on other parameters ( $\pi$ ,  $\sigma^2$ , or  $\tau$ ).

# HIERARCHICAL MODEL SPECIFICATION

Conceptually, can think of process in 3 steps:

1. **Estimate exposure effects ( $\beta_j$ ).**
2. **Rgress exposure effects on the characteristics of those exposures ( $z$ ),** estimating the *characteristic-specific effects* of each of the exposures ( $\pi$ ).
3. Update estimate of exposure effects given step 2.

# HIERARCHICAL MODEL SPECIFICATION

Consider: effect of set of exposures ( $\mathbf{x}$ ) on disease state ( $y$ ):

- **First stage** model:

$$g(E[Y|\mathbf{x}, \mathbf{w}]) = \alpha + \sum_j x_j \beta_j + \sum_I w_I \gamma_I \quad (1)$$

---

\*Confusingly, some authors use  $\tau$  to refer to variance.

# HIERARCHICAL MODEL SPECIFICATION

Consider: effect of set of exposures ( $\mathbf{x}$ ) on disease state ( $y$ ):

- First stage model:

$$g(E[Y|\mathbf{x}, \mathbf{w}]) = \alpha + \sum_j x_j \beta_j + \sum_l w_l \gamma_l \quad (1)$$

confounders

---

\*Confusingly, some authors use  $\tau$  to refer to variance.

# HIERARCHICAL MODEL SPECIFICATION

Consider: effect of set of exposures ( $\mathbf{x}$ ) on disease state ( $y$ ):

- First stage model:

$$g(E[Y|\mathbf{x}, \mathbf{w}]) = \alpha + \sum_j x_j \beta_j + \sum_l w_l \gamma_l \quad (1)$$

↑  
exposures (foods,  
chemicals, etc...)      confounders

---

\*Confusingly, some authors use  $\tau$  to refer to variance.

# HIERARCHICAL MODEL SPECIFICATION

Consider: effect of set of exposures ( $\mathbf{x}$ ) on disease state ( $y$ ):

- **First stage model:**

$$g(E[Y|\mathbf{x}, \mathbf{w}]) = \alpha + \sum_j x_j \beta_j + \sum_l w_l \gamma_l \quad (1)$$

exposures (foods,  
chemicals, etc...)      confounders

- **Second stage model:**

$$\beta_j = \sum_k z_{jk} \pi_k + \delta_j \quad (2)$$

$$\delta_j \sim N(0, \sigma^2)$$

for  $j = 1, \dots, J$  exposures. Here  $\sigma^2$  represents residual *variance* (can be equivalently specified with precision  $\tau = 1/\sigma^2$ ).\*

\*Confusingly, some authors use  $\tau$  to refer to variance.

# HIERARCHICAL MODEL SPECIFICATION

Consider: effect of set of exposures ( $\mathbf{x}$ ) on disease state ( $y$ ):

- **First stage model:**

$$g(E[Y|\mathbf{x}, \mathbf{w}]) = \alpha + \sum_j x_j \beta_j + \sum_l w_l \gamma_l \quad (1)$$

exposures (foods,  
chemicals, etc...)      confounders

- **Second stage model:**

characteristics of exposures  
(nutrients, chemical classes, etc...)

$$\begin{aligned} \beta_j &= \sum_k z_{jk} \pi_k + \delta_j \\ \delta_j &\sim N(0, \sigma^2) \end{aligned} \quad (2)$$

for  $j = 1, \dots, J$  exposures. Here  $\sigma^2$  represents residual *variance* (can be equivalently specified with precision  $\tau = 1/\sigma^2$ ).\*

\*Confusingly, some authors use  $\tau$  to refer to variance.

# HIERARCHICAL MODEL SPECIFICATION

Consider: effect of set of exposures ( $\mathbf{x}$ ) on disease state ( $y$ ):

- **First stage model:**

$$g(E[Y|\mathbf{x}, \mathbf{w}]) = \alpha + \sum_j x_j \beta_j + \sum_l w_l \gamma_l \quad (1)$$

exposures (foods,  
chemicals, etc...)      confounders

- **Second stage model:**

characteristics of exposures  
(nutrients, chemical classes, etc...)

$$\begin{aligned} \beta_j &= \sum_k z_{jk} \pi_k + \delta_j && \text{random (residual) food-specific effect} \\ \delta_j &\sim N(0, \sigma^2) \end{aligned} \quad (2)$$

for  $j = 1, \dots, J$  exposures. Here  $\sigma^2$  represents residual *variance* (can be equivalently specified with precision  $\tau = 1/\sigma^2$ ).\*

\*Confusingly, some authors use  $\tau$  to refer to variance.

# HIERARCHICAL MODEL SPECIFICATION

Characteristics drive the effects of the exposures.

- 2<sup>nd</sup> stage model essentially regresses 1<sup>st</sup> stage exposure effects ( $\beta^s$ ) on a set of predictors ( $z$ ).
  - e.g. in Witte *et al* (2000, *Epidemiology*),  $z$  is given by *nutrients* thought to drive relationship between individual foods and breast cancer risk:

TABLE 2. Second-Stage Design Matrix for Multilevel Model

Food	Glucosinolates (mg)*	Carotenoids ( $\mu\text{g}$ )†			Vitamin E ( $\mu\text{g}$ )‡	Organosulfate (mg)‡	Fiber (gm)‡
		Alpha-Carotene	Beta-Carotene	Lutein			
Cauliflower	48	0	8	33	0	95	234
Broccoli	74	1	1,300	1,800	0	660	351
Brussels sprouts	192	6	480	1,300	0	557	547
Red cabbage	66	0.7	32	58	0	1,740	204
White cabbage	63	0	80	150	0	1,720	241
Savoy cabbage	87	0	80	150	0	2,560	174
Sauerkraut	73	0	80	150	0	153	1,630
Allium vegetable	0	0	481	736	0	300	236
Tomato	0	0	567	109	3,780	869	417
Pizza	0	0	170	33	650	1,280	303
Tomato salad	0	0	263	170	1,550	2,740	557
Strawberries	0	2	9	31	0	120	1,000

Values correspond to the amount of relevant constituents in 100 gm of given foods (if foods are combined, then a weighted average is used here, where the weight reflects the relative consumption of each food).

\* From Fenwick *et al.*<sup>31</sup>.

† From database created by the National Cancer Institute.<sup>32</sup>

‡ From database of German foods developed in Berlin.<sup>33</sup>

# HIERARCHICAL MODEL SPECIFICATION

## Food-level effects from 2<sup>nd</sup> stage model:

- e.g. the effect of a 100g increase in *cauliflower* consumption:

$$\begin{aligned}\beta_1 = & 48\pi_1 + 0\pi_2 + 8\pi_3 + 33\pi_4 + 0\pi_5 + \\& 95\pi_6 + 234\pi_7 + 2.87\pi_8 + \delta_1\end{aligned}$$

- Where  $\pi_1$  represents the influence of 1mg increase in glucosinolate intake on risk of breast cancer,  $\pi_2$  represents the influence of 1 $\mu$ g increase in  $\alpha$ -Carotene intake on breast cancer, etc...
  - **Hierarchical** because  $\beta$  depends on  $\pi$  and  $\sigma^2$ .
  - $\pi$  is estimated in the modeling process.
- Generally,  $\mathbf{z}$  (2<sup>nd</sup> stage predictors) can include indicators of chemical classes, metabolic pathways, etc...

# HIERARCHICAL MODEL SPECIFICATION

The error term (residual)  $\delta_j$  represents the influence of *other factors not considered* in the  $\mathbf{z}$  matrix.

- In mixed model terminology:  $\delta_j$  is the *random effect* of food  $j$ .

$$\delta_j \sim N(0, \sigma^2) \quad \text{for } j = 1, \dots, J$$

- Can estimate residual *variance*  $\sigma^2$  from data, or specify to yield plausible range of effects.\*
- Authors (Greenland *et al*) suggest set  $\sigma^2 = 0.35^2 = 0.1225$  (SG typically denotes variance as  $\tau$ ).
  - Corresponds to 95% *prior* intervals for *ratio* effects within 4-fold range of point estimate (e.g. 0.5-2.0 if OR=1).

---

\*This yields good CI coverage and MSE; see Greenland S. Methods for epidemiologic analysis of multiple exposures: a review and comparative study of maximum likelihood, preliminary-testing, and empirical-Bayes regression. *Stat Med* 1993; 12:717-736.

# INSIGHT FROM MIXED MODEL

Another way to think about the hierarchical model; substituting (2) into (1):

$$g [E(Y|\mathbf{x}, \mathbf{w}, \mathbf{z}, \boldsymbol{\delta})] = a + \sum_k \left( \sum_j x_j z_{jk} \right) \pi_k + \sum_j x_j \delta_j + \mathbf{w}\gamma$$
$$\delta_j \sim N(0, \sigma^2)$$

# INSIGHT FROM MIXED MODEL

Another way to think about the hierarchical model; substituting (2) into (1):

$$g [E(Y|\mathbf{x}, \mathbf{w}, \mathbf{z}, \boldsymbol{\delta})] = a + \sum_k \underbrace{\left( \sum_j x_j z_{jk} \right)}_{\text{Total exposure to characteristic } k} \pi_k + \sum_j x_j \delta_j + \mathbf{w}\gamma$$
$$\delta_j \sim N(0, \sigma^2)$$

# INSIGHT FROM MIXED MODEL

Another way to think about the hierarchical model; substituting (2) into (1):

$$g [E(Y|\mathbf{x}, \mathbf{w}, \mathbf{z}, \boldsymbol{\delta})] = a + \sum_k \underbrace{\left( \sum_j x_j z_{jk} \right)}_{\text{Total exposure to characteristic } k} \pi_k + \sum_j x_j \delta_j + \mathbf{w}\gamma$$

$\delta_j \sim N(0, \sigma^2)$       residual effects of individual exposures

# INSIGHT FROM MIXED MODEL

Another way to think about the hierarchical model; substituting (2) into (1):

$$g [E(Y|\mathbf{x}, \mathbf{w}, \mathbf{z}, \boldsymbol{\delta})] = a + \sum_k \underbrace{\left( \sum_j x_j z_{jk} \right)}_{\text{Total exposure to characteristic } k} \pi_k + \sum_j x_j \delta_j + \mathbf{w}\gamma$$

$\delta_j \sim N(0, \sigma^2)$       residual effects of individual exposures

- $\pi_k$  is the *fixed effect* of **characteristic**  $k$  on disease risk.
- $\delta_j$  is the *random effect* of **exposure**  $j$  on disease risk.
  - Factors that are not related to constituents represented by characteristics ( $\mathbf{z}$ ).

# INTERPRETATION

- **1<sup>st</sup> stage effect estimates:** effect of exposure  $j$  on outcome  $Y$ .

$$\hat{\beta}_j = \sum_k z_{jk} \hat{\pi}_k + \hat{\delta}_j$$

- Does not force similar exposures to have identical effects—only *similar* (up to characteristics/information in 2<sup>nd</sup> stage model).
- Dissimilarity accounted for in residual error ( $\delta$ ).
- **2<sup>nd</sup> stage effect estimates** ( $\pi$ ): associations for individual characteristics on the outcome.

# HIERARCHICAL MODEL RESULTS

Witte et al. (2000) Diet and breast cancer risk:

- Compared to MLE with all 12 foods included simultaneously to hierarchical model.
- Point estimates:** extreme associations attenuated (e.g. white cabbage, pizza).
- Uncertainty intervals** tend to be more precise.

TABLE 1. Odds Ratios and 95% Confidence Intervals from Conventional Maximum-Likelihood and Multilevel (Semi-Bayes) Analyses of Data

Food*	Maximum-Likelihood Logistic Regression		Multilevel Model†			
	OR	95% CI	GLIMMIX		Two-Step Procedure	
Cauliflower	1.0	0.83–1.3	1.0	0.84–1.3	1.0	0.84–1.3
Broccoli	0.83	0.29–2.4	0.75	0.28–2.0	0.75	0.28–2.0
Brussels sprouts	6.1	1.5–26	5.8	1.5–23	5.6	1.4–22
Red cabbage	1.3	0.53–3.3	1.3	0.62–2.7	1.3	0.62–2.7
White cabbage	2.5	0.75–8.4	1.2	0.59–2.6	1.2	0.60–2.6
Savoy cabbage	0.58	0.16–2.1	0.91	0.37–2.2	0.91	0.37–2.3
Sauerkraut	0.26	0.08–0.88	0.31	0.10–0.96	0.32	0.10–1.0
Allium vegetables	0.69	0.44–1.1	0.73	0.50–1.1	0.74	0.49–1.1
Tomato	1.1	0.75–1.6	1.1	0.74–1.5	1.1	0.74–1.5
Pizza	0.48	0.05–4.5	0.77	0.32–1.8	0.78	0.32–1.9
Tomato salad	0.54	0.13–2.2	0.57	0.19–1.7	0.58	0.19–1.8
Strawberries	0.89	0.30–2.6	0.87	0.31–2.4	0.87	0.31–2.4

\* Results correspond to a 100-gm-per-week increase (vs none) in the foods.

† Semi-Bayes,  $\tau_i = 0.35$  for each food.

# ESTIMATION

- **Fully Bayesian** framework:
  - Specify priors on all free parameters (non-informative for 1<sup>st</sup> stage covariate/2<sup>nd</sup> stage effects).
  - Characterize posterior distribution of model parameters with MCMC sampling (Stan, JAGS, Nimble, OpenBUGS, proc mcmc [SAS], bayesmh [Stata]).
- **Semi-Bayesian** framework (priors/penalties on only some parameters):
  - 2-stage least squares approach.\*
  - Mixed models (Witte and Greenland. *Epidemiology*. 2000).
  - Penalized maximum likelihood.

---

\*Greenland S. Second-stage least squares versus penalized quasi-likelihood for fitting hierarchical models in epidemiologic analyses. *Statistics in Medicine* (1997). Witte JS, Greenland S, and Kim LL. Software for hierarchical modeling of epidemiologic data. *Epidemiology* (1998).

# ESTIMATION

Estimation in a fully-Bayesian framework (with MCMC sampling from the posterior) is a trivial extension on previous code for regression modeling.

- Only requires specifying an additional (linear) regression model for  $\beta$  as a function of  $\mathbf{z}$  and  $\pi$ .
- Also, explicit priors needed for other model parameters ( $a, \gamma, \pi$  from previous specification).

# ESTIMATION: JAGS MODEL CODE

```
1 hierarchical.model <- function() {  
2     for (i in 1:N) {### First stage model (likelihood):  
3         logit(p[i]) <- a + inprod(X[i,1:Nx],b[1:Nx]) +  
4                         inprod(W[i,(1:Nw)],g[1:Nw]);  
5         Y[i] ~ dbin(p[i],1);  
6     }  
7     for (j in 1:Nx) {### Second stage model (prior on beta--linear model):  
8         mub[j] <- inprod(Z[j,1:Nz],pi[1:Nz]);  
9         b[j] ~ dnorm(mub[j], taub); # taub is PRECISION (1/prior variance)  
10        expb[j] <- exp(b[j]); # Calculate Odds Ratio for each covariate effect  
11    }  
12    # PRIORS ON PARAMETERS (all non-informative with precision (tau)=1/1000  
13    for (l in 1:Nz) {pi[l] ~ dnorm(0, 0.001); expp[l] <- exp(pi[l]);} # 2nd stage  
14    a ~ dnorm(0,0.001); # 1st stage intercept  
15    for (l in 1:Nw) { g[l] ~ dnorm(0, 0.001); } # confounder effects  
16 }
```

# ESTIMATION: JAGS CODE

## Notes:

- **Notation:**  $\text{inprod}(X[i, 1:Nx], b[1:Nx]) = \sum_{j=1}^{Nx} x_{i,j}\beta_j$
- **1<sup>st</sup> stage model** as previously presented:
  - Priors for  $\alpha$  and  $\gamma$  as usual (non-informative).
  - Prior for  $\beta$  given by 2<sup>nd</sup> stage model.
- **2<sup>nd</sup> stage model:**
  - Essentially linear regression, treating  $\beta$  from 1<sup>st</sup> stage model as **outcomes**.
  - Note that 2<sup>nd</sup> stage covariate matrix (predictors,  $\mathbf{z}$ ) is of different dimension than 1<sup>st</sup> stage covariates ( $\mathbf{x}$ ,  $\mathbf{w}$ ):
    - $\mathbf{z}$  has same number of *rows* (observations) as number of exposures (*columns* of  $\mathbf{x}$ ).
    - *Columns* of  $\mathbf{z}$  correspond to *characteristics* of exposures (could be 1 or more).
    - Prior on  $\pi$  is non-informative (although could be informative if one has information on  $\mathbf{z}$  (characteristic)-specific effects).

# APPLICATION: PESTICIDES AND NON-HODGKIN'S LYMPHOMA

## APPLICATION: PESTICIDES AND NHL

De Roos *et al.* *Occupational and Environmental Medicine* (2003): Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men.

- Pooled analysis of 3 case-control studies from US National Cancer Institute.
- Analysis included 650 cases/1933 controls.
- Analyses for individual effects (simultaneously adjusted) as well as *joint effects* (combinations of pesticides):
  - Standard multivariable logistic regression (no 2<sup>nd</sup> stage model).
  - Bayesian hierarchical model.

# APPLICATION: PESTICIDES AND NHL

$$\text{logit} [E(Y|\mathbf{x}, \mathbf{w})] = a + \sum_j x_j \beta_j + \sum_k w_k \gamma$$

$$\beta_j = \sum_k z_{jk} \pi_k + \delta_j$$

- **x=exposure to 47 pesticides** for each subject.
- **z=9 characteristics** of each of the 47 pesticides.

**Table 1** Second-level matrix for hierarchical regression analysis, showing values of "prior covariates" for each pesticide of interest\*†

Pesticides	Insecticides	Organochlorines	Organophosphates	Carbamates	Phenoxy-acetic acids	Triazines	Amides	Benzoic acids	Carcinogenic probability
<b>Insecticides</b>									
Aldrin	1	1	0	0	0	0	0	0	0.6
Bufencarb	1	0	0	1	0	0	0	0	0.3
Carbaryl	1	0	0	1	0	0	0	0	0.3
Carbofuran	1	0	0	1	0	0	0	0	0.3
Chlordane	1	1	0	0	0	0	0	0	0.8
Copper acetoarsenite*	1	0	0	0	0	0	0	0	1.0
Coumaphos	1	0	1	0	0	0	0	0	0.3
DDT	1	1	0	0	0	0	0	0	0.8

(truncated for space)

# PESTICIDE-SPECIFIC ESTIMATES

## Insecticide exposures: (partial list)

- Note generally more narrow CIs from hierarchical regression.
- Several effect estimates (Coumaphos, Dieldrin) notably attenuated.

Pesticides	Exposed [n (%)]		Logistic regression OR (95% CL)†	Hierarchical regression OR (95% CL)
	Cases (n=650)	Controls (n=1933)		
<b>Insecticides</b>				
Aldrin	47 (7.2%)	115 (5.9%)	0.5 (0.3 to 0.9)	0.6 (0.4 to 1.0)
Bufencarb‡	6 (0.9%)	12 (0.6%)	1.1 (0.3 to 3.7)	1.0 (0.4 to 2.3)
Carbaryl	30 (4.6%)	57 (2.9%)	1.0 (0.5 to 1.9)	1.1 (0.6 to 1.9)
Carbofuran	41 (6.3%)	96 (5.0%)	0.9 (0.5 to 1.6)	1.0 (0.6 to 1.7)
Chlordane	39 (6.0%)	65 (3.4%)	1.5 (0.8 to 2.6)	1.3 (0.8 to 2.1)
Copper acetoarsenite	41 (6.3%)	68 (3.5%)	1.4 (0.9 to 2.3)	1.4 (0.9 to 2.1)
Coumaphos	15 (2.3%)	22 (1.1%)	2.4 (1.0 to 5.8)	1.7 (0.9 to 3.3)
DDT	98 (15.1%)	226 (11.7%)	1.0 (0.7 to 1.3)	1.0 (0.7 to 1.3)
Diazinon	40 (6.1%)	62 (3.2%)	1.9 (1.1 to 3.6)	1.7 (1.0 to 2.8)
Dichlorvos	16 (2.5%)	37 (1.9%)	0.9 (0.4 to 2.0)	0.9 (0.5 to 1.7)
Dieldrin	21 (3.2%)	39 (2.0%)	1.8 (0.8 to 3.9)	1.4 (0.8 to 2.6)
Dimethoate‡	5 (0.8%)	11 (0.6%)	1.2 (0.3 to 5.3)	1.2 (0.5 to 2.8)

# PESTICIDE-SPECIFIC ESTIMATES (SELECTED)

## Herbicide exposures: (partial list)

- Similar patterns.
- Estimates for glyphosate and sodium chlorate attenuated to more reasonable magnitudes.

Herbicides				
Alachlor	68 (10.5%)	152 (7.9%)	1.1 (0.7 to 1.8)	1.0 (0.6 to 1.6)
Atrazine	90 (13.8%)	185 (9.6%)	1.6 (1.1 to 2.5)	1.5 (1.0 to 2.2)
Bentazon	22 (3.4%)	58 (3.0%)	0.7 (0.3 to 1.5)	0.8 (0.4 to 1.4)
Butylate	28 (4.3%)	56 (2.9%)	1.2 (0.6 to 2.3)	1.2 (0.7 to 2.0)
Chloramben	34 (5.2%)	81 (4.2%)	0.9 (0.5 to 1.6)	0.9 (0.5 to 1.5)
Cyanazine	37 (5.7%)	96 (5.0%)	0.6 (0.3 to 1.0)	0.6 (0.4 to 1.1)
2,4-D	123 (18.9%)	314 (16.2%)	0.8 (0.6 to 1.1)	0.9 (0.6 to 1.2)
Dicamba	39 (6.0%)	79 (4.1%)	1.2 (0.6 to 2.3)	1.2 (0.7 to 2.1)
EPTC + protectant	13 (2.0%)	29 (1.5%)	1.2 (0.5 to 3.1)	1.1 (0.5 to 2.3)
Glyphosate	36 (5.5%)	61 (3.2%)	2.1 (1.1 to 4.0)	1.6 (0.9 to 2.8)
Linuron	5 (0.8%)	22 (1.1%)	0.3 (0.1 to 1.2)	0.5 (0.2 to 1.2)
MCPA	8 (1.2%)	16 (0.8%)	1.0 (0.4 to 2.6)	0.9 (0.4 to 2.0)
Metolachlor	13 (2.0%)	37 (1.9%)	0.7 (0.3 to 1.6)	0.7 (0.4 to 1.5)
Metribuzen	20 (3.1%)	53 (2.7%)	0.8 (0.4 to 1.7)	0.8 (0.4 to 1.5)
Paraquat‡	2 (0.3%)	15 (0.8%)	0.1 (0.02 to 0.7)	0.5 (0.2 to 1.2)
Propachlor	20 (3.1%)	50 (2.6%)	1.0 (0.5 to 2.0)	1.0 (0.6 to 1.9)
Sodium chlorate‡	8 (1.2%)	7 (0.4%)	4.1 (1.3 to 13.6)	1.8 (0.8 to 4.1)

# PESTICIDE-SPECIFIC ESTIMATES (SELECTED)

## Joint effects (partial):

- Similar benefit for effects on *combinations* of pesticides.

Individual and joint pesticide exposures	Exposed [n (%)]		Logistic regression OR (95% CL)†	Hierarchical regression OR (95% CL)
	Cases (n=650)	Controls (n=1933)		
Alachlor and atrazine				
Neither	545	1695	1.0	1.0
Alachlor only	15 (2.3%)	53 (2.7%)	0.7 (0.3 to 1.3)	0.7 (0.4 to 1.3)
Atrazine only	37 (5.7%)	86 (4.5%)	1.3 (0.8 to 2.1)	1.2 (0.8 to 1.8)
Both	53 (8.2%)	99 (5.1%)	2.1 (1.1 to 3.9)	1.6 (1.0 to 2.7)
Atrazine and dicamba				
Neither	552	1729	1.0	1.0
Atrazine only	59 (9.1%)	125 (6.5%)	1.5 (1.0 to 2.4)	1.4 (0.9 to 2.0)
Dicamba only	8 (1.2%)	19 (1.0%)	0.9 (0.3 to 2.6)	1.0 (0.5 to 2.0)
Both	31 (4.8%)	60 (3.1%)	2.1 (1.0 to 4.7)	1.6 (0.9 to 2.9)

# MORE RECENT APPLICATIONS

Hyland et al. (2021) *Environmental Epidemiology*. 5:e150.

Original Research Article



ENVIRONMENTAL  
EPIDEMIOLOGY

OPEN

## Associations between pesticide mixtures applied near home during pregnancy and early childhood with adolescent behavioral and emotional problems in the CHAMACOS study

Carly Hyland<sup>a</sup>, Patrick T. Bradshaw<sup>b</sup>, Robert B. Gunier<sup>a</sup>, Ana M. Mora<sup>a,c</sup>, Katherine Kogut<sup>a</sup>, Julianna Deardorff<sup>a</sup>, Sharon K. Sagiv<sup>a,b</sup>, Asa Bradman<sup>a,d</sup>, Brenda Eskenazi<sup>a,\*</sup>

Hamra et al. (2019) *Epidemiology* 30:418-426.

## Prenatal Exposure to Endocrine-disrupting Chemicals in Relation to Autism Spectrum Disorder and Intellectual Disability

Ghassan B. Hamra,<sup>a</sup> Kristen Lyall,<sup>b</sup> Gayle C. Windham,<sup>c</sup> Antonia M. Calafat,<sup>d</sup> Andreas Sjödin,<sup>d</sup> Heather Volk,<sup>e</sup> and Lisa A. Croen<sup>f</sup>

# FINAL THOUGHTS

# ADVANTAGES

- Encourages **engagement with subject matter**.
- **Inference remains on relevant unit of exposure.**
- **Improved precision** compared to standard multi-exposure modeling (Witte and Greenland. *Stat Med*. 1996):
  - Narrow interval estimates.
  - Lower MSE.
- Shrinkage estimators assuage issues around **multiple comparisons**.
  - Less likely to observe spurious non-null associations.
- Fully Bayesian framework  $\Rightarrow$  other benefits (informative priors, bias analysis, missing data, etc...).

# CHALLENGES

- **Scaling exposures** so that effects are exchangeable.
- 1<sup>st</sup> stage exposure variables (**x**) should be scaled to reflect similar magnitudes of exposure.
- **Serving size** for foods (e.g. per 100g).
  - Cf. treatment of foods for a priori dietary index scoring.
- Other common approaches:
  - Dichotomize?
  - Standardize? (See applications papers in reference list)
  - Probably no ideal solution (see Greenland. *Stat Med.* 1992).
- Parametric assumptions around 2<sup>nd</sup> stage residual term (see MacLehose (2007) *Epidemiology* for one solution).
- BHM is *one* approach to multiple-exposure modeling; others may have advantages for particular purposes (see review by Hamra and Buckley (2018) *Curr Epidemiol Rep*).

# REFERENCES

# FOUNDATIONAL READINGS

Greenland, S. Section titled: Hierarchical (multilevel) regression in Chapter 21: Introduction to regression modeling. *Modern epidemiology* (eds Kenneth J. Rothman, Sander Greenland, & Timothy L. Lash), (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008). (p435-439)

- Greenland S. Methods for epidemiologic analyses of multiple exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. *Stat Med* 12, 717-736 (1993).
- Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. *Environ Health Perspect* 102 Suppl 8, 33-39, doi:10.1289/ehp.94102s833 (1994).
- Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 29, 158-167 (2000).
- Hamra G. B. & Buckley J. P. Environmental exposure mixtures: questions and methods to address them. *Curr Epidemiol Rep* 5, 160-165, doi:10.1007/s40471-018-0145-0 (2018).
- MacLehose R. F., Dunson D. B., Herring A. H. & Hoppin J. A. Bayesian methods for highly correlated exposure data. *Epidemiology* 18, 199-207, doi:10.1097/01.ede.0000256320.30737.c0 (2007).
- Witte J. S. & Greenland S. Simulation study of hierarchical regression. *Stat Med* 15, 1161-1170 (1996).
- Witte J. S., Greenland S., Kim L. L. & Arab L. Multilevel modeling in epidemiology with GLIMMIX. *Epidemiology* 11, 684-688 (2000).
- Witte J. S. Genetic analysis with hierarchical models. *Genet Epidemiol* 14, 1137-1142 (1997).

## APPLICATIONS: ENVIRONMENTAL

- De Roos A. J., Poole C., Teschke K. & Olshan A. F. An application of hierarchical regression in the investigation of multiple paternal occupational exposures and neuroblastoma in offspring. *Am J Ind Med* 39, 477-486 (2001).
- De Roos A, Zahm SH, Cantor KP, Weisenburger DD, Holmes FF, Burmeister LF, Blair A. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. *Occupational and Environmental Medicine*. 60.9 (2003).
- Braun J. M. et al. Gestational exposure to endocrine-disrupting chemicals and reciprocal social, repetitive, and stereotypic behaviors in 4- and 5-year-old children: the HOME study. *Environ Health Perspect* 122, 513-520, doi:10.1289/ehp.1307261 (2014).
- White A. J. et al. Exposure to multiple sources of polycyclic aromatic hydrocarbons and breast cancer incidence. *Environ Int* 89-90, 185-192, doi:10.1016/j.envint.2016.02.009 (2016).
- Hamra, Ghassan B., et al. Prenatal Exposure to Endocrine-disrupting Chemicals in Relation to Autism Spectrum Disorder and Intellectual Disability. *Epidemiology* 30.3 (2019): 418-426.

## APPLICATIONS: DIET

- Witte J. S., Greenland S., Haile R. W. & Bird C. L. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology* 5, 612-621 (1994).
- Bradshaw P. T. et al. Nutrient pathways and breast cancer risk: the Long Island Breast Cancer Study Project. *Nutr Cancer* 65, 345-354, doi:10.1080/01635581.2013.767364 (2013).
- Carmichael S. L., Witte J. S. & Shaw G. M. Nutrient pathways and neural tube defects: a semi-Bayesian hierarchical analysis. *Epidemiology* 20, 67-73, doi:10.1097/EDE.0b013e31818f6375 (2009).

## APPLICATIONS: GENETIC

- Hung R. J. et al. Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. *Cancer Epidemiol Biomarkers Prev* 13, 1013-1021 (2004).
- Khankari N. K. et al. Genetic variation in multiple biologic pathways, flavonoid intake, and breast cancer. *Cancer Causes Control* 25, 215-226, doi:10.1007/s10552-013-0324-8 (2014).
- Engel S. A. et al. Risk of spontaneous preterm birth is associated with common proinflammatory cytokine polymorphisms. *Epidemiology* 16, 469-477 (2005).

# THANK YOU

Thank you to organizers of ALEXS-2021 for inviting me.

Patrick T. Bradshaw, Ph.D.

Division of Epidemiology, School of Public Health

University of California, Berkeley

✉ pbradshaw@berkeley.edu

/github.com/patrickbradshaw/ALEXS2021-BHM

## **ADDITIONAL APPLICATION: NUTRIENT PATHWAYS**

# APPLICATION: NUTRIENT PATHWAYS

Bradshaw *et al.* *Nutrition and Cancer* (2013).

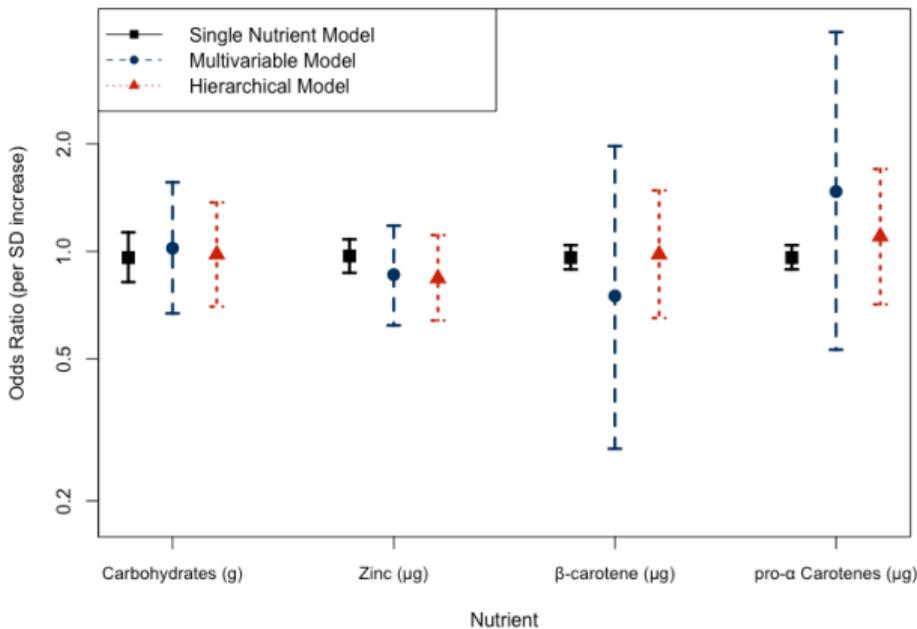
- Several nutrients have been implicated in breast cancer risk.
  - Which diet-influenced *pathways* are most relevant?
- Case-control analysis from Long Island Breast Cancer Study.
  - $x$ =intake of 33 **nutrients** for each subject (scaled to 1-SD units).
  - $z$ =indicators of 4 relevant **biological pathways** for each nutrient.

Nutrient	Nutrient Pathway			
	One-carbon Metabolism	Oxidative Stress	Glycemic Control	Phytoestrogen
Carbohydrates	0	0	-1	0
Calcium	0	0	1	0
Magnesium	0	0	1	0
Zinc	1	1	1	0
$\alpha$ -carotene	0	1	0	0
$\beta$ -carotene	0	1	0	0
Iron	0	-1	0	0
:	:	:	:	:

# NUTRIENT-SPECIFIC ESTIMATES (SELECTED)

$$\hat{\beta}_j = \sum_k z_{jk} \hat{\pi}_k + \hat{\delta}_j$$

Nutrient Intake and Breast Cancer Risk



## PATHWAY-SPECIFIC ESTIMATES

$$\hat{\beta}_j = \sum_k z_{jk} \hat{\pi}_k + \hat{\delta}_j$$

Nutrient Pathway	OR <sup>a</sup>	95% CI
One-carbon	1.06	(0.87, 1.31)
Oxidative stress	0.97	(0.78, 1.19)
Glycemic control	0.86	(0.62, 1.21)
Phytoestrogen	1.00	(0.76, 1.33)

A 1-SD increase in intake of a nutrient related to glycemic control resulted in a 14% reduction in odds of breast cancer.