

G **T** **S** **G** **T**

5 **E** **5** **E**

N **O** **N** **O**

O **S** **O** **S**

G **T** **G** **T**

5 **E** **5** **E**

N **O** **N** **O**

O **S** **O** **S**

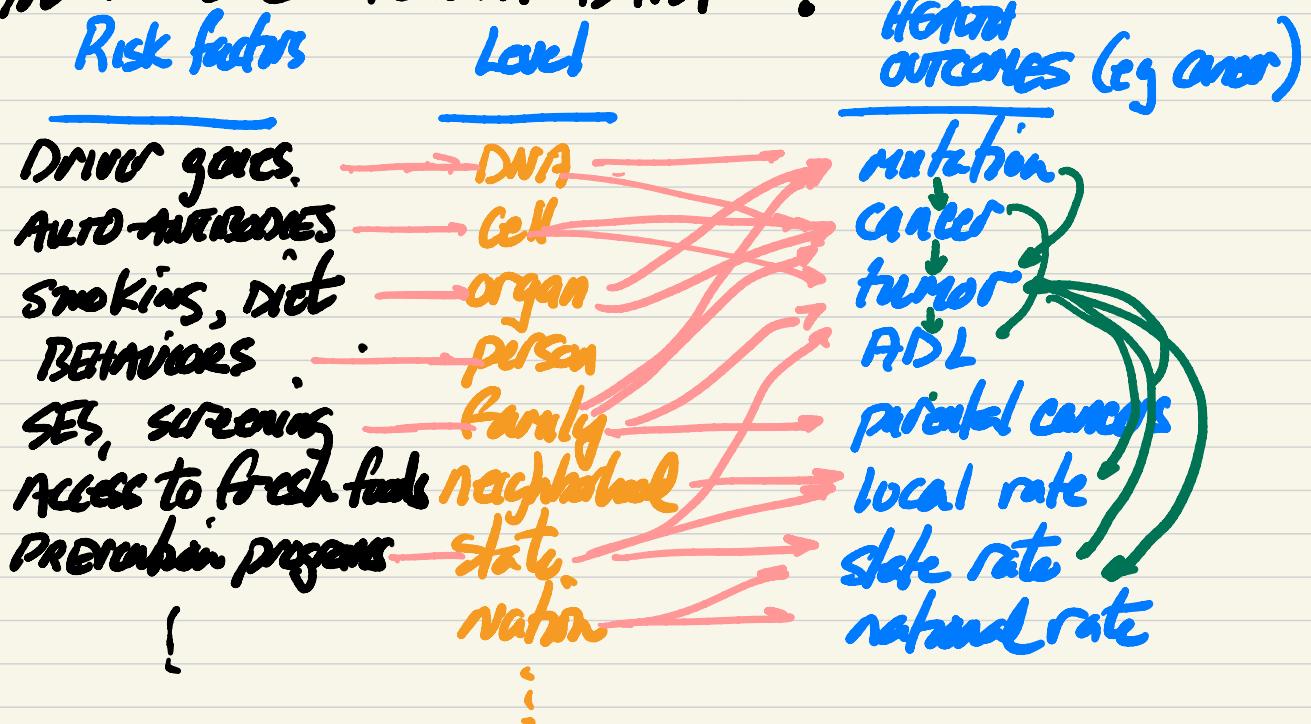
G **T** **G** **T**

5 **E** **5** **E**

Mult-Level Models (MLMS) 2021

Final Questions for
Discussion

1. WHAT IS THE MULTI-LEVEL (CAUSAL) MODEL FOR PUBLIC HEALTH DATA ?



Key ideas

- risk factors and outcomes occur and different spatial + temporal scales that interact with one another
- interventions targeted to levels : eg smoking : smoker education or patch -vs- no smoking in public/private buildings
- decompose evidence about relationship $X \rightarrow Y$ into different level specific components + combine when appropriate

- clustered responses ($y_{ij}, y_{ij'}$) are likely correlated; correlation must be modelled to obtain:
 - valid inferences
 - efficient estimators
 - robustness to missing data patterns
- there are heterogeneities in covariate effects among clusters at each level that give rise to the corrections.
of interest

2. WRITE and INTERPRET an interesting ("rich") 2-LEVEL MODEL TO ADDRESS THE QUESTION: does an 11-year olds' family SES affect his or her performance on a reading comprehension test and is the SES effect the same for boys and girls?

Y_{ij} - reading test score for child $j = 1, \dots, n_i$ in School $i = 1, \dots, m$

F_{ij} = 1 if female; 0 if male

S_{ij} = SES level

$$Y_{ij} = \beta_{0i} + \beta_1(S_{ij} - \bar{S}_i) + \beta_2 F_{ij} + \beta_3(S_{ij} - \bar{S}_i)F_{ij} + \epsilon_{ij}$$

$$\begin{pmatrix} \beta_{0i} \\ \beta_1 \\ \beta_2 \end{pmatrix} \sim G\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, D\right), \quad \epsilon_{ij} \stackrel{iid}{\sim} G(0, \sigma^2)$$

(β_{0i}) - (intercept) that define the linear dependence
 (β_{1i}) - (S slope) of reading on SES within all boys,
 β_2 - average difference in reading score comparing
girls to boys of the same SES from same
school *within school*

β_3 - the increase in the SES effect on average
reading comparing an all girls to all boys
school

$D = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$: $d_{11} = \text{Var}(\beta_{0i}) \Rightarrow \pm 2\sqrt{d_{11}}$ will include
 $\approx 95\%$ of intercepts; $d_{22} = \text{Var}(\beta_{1i}) \Rightarrow$
 $\pm 2\sqrt{d_{22}}$ will include 95% of Y-S slopes
around their mean size that depends on \bar{F}_{12}

σ^2 - residual variance ($\text{Var} \epsilon_{ij}$)

3. Continuing from (2), does the contextual effect of the average School SES differ for boys and girls?

$$Y_{ij} = \beta_{0i} + \beta_1(S_{ij} - \bar{S}_{..}) + \beta_2(\bar{S}_i - \bar{S}_{..}) + \beta_3 F_{ij} + \beta_4(Z_{ij} - \bar{Z}_{..}) F_{ij} + \varepsilon_{ij}$$

♀ Contextual effect

Yes $\beta_2 + \beta_4$

No β_2

4. Consider a 3-level design in which children (k) within families (j) within villages (i) have lower respiratory infection ($Y_{ijk}=1$) or not ($Y_{ijk}=0$). GRESSTIMATE and Interpret the log odds ratio ($Y_{ijk}, Y_{i'j'k'}$) when :

$$4.1 \quad l = l', j = j', k \neq k'$$

$$\frac{\log OR(Y_{ijk}, Y_{i'j'k'})}{\alpha_0 + \alpha_1}$$

$$4.2 \quad l = l', j \neq j', k \neq k'$$

$$\alpha_0$$

$$4.3 \quad l \neq l', j \neq j', k \neq k'$$

$$0$$

$$\log OR(Y_{ijk}, Y_{i'j'k'}) = \alpha_0 \mathbb{1}_{\{l=l'\}} + \alpha_1 \mathbb{1}_{\{l=l', j=j'\}}$$

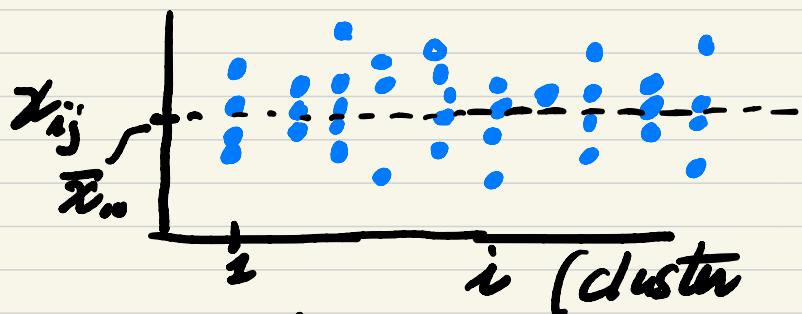
5. Consider the 2-level model:

$$Y_{ij} = \beta_{0i} + \beta_{1W}(X_{ij} - \bar{X}_{i\cdot}) + \beta_{2B}\bar{X}_{i\cdot} + \varepsilon_{ij}$$

$$\beta_{0i} = \beta_0 + b_i; \quad b_i \stackrel{iid}{\sim} G(0, \tau^2), \quad \varepsilon_{ij} \stackrel{iid}{\sim} G(0, \sigma^2)$$

$$\text{Cov}(b_i, \varepsilon_{ij}) = 0, \text{ for all } i, j$$

and the data set shown below



$$\frac{\text{Var}(\hat{\beta}_{1W})}{\text{Explained}} \leq \frac{\text{Var}(\hat{\beta}_{1B})}{(a)} \cdot \begin{matrix} \leq, =, \geq, \\ (a) \end{matrix} \begin{matrix} \geq, \\ (b) \end{matrix} \begin{matrix} \leq, =, \geq, \\ (c) \end{matrix} \begin{matrix} \geq, \\ (d) \end{matrix}$$

6. If $\hat{\beta}_{1w} \neq \hat{\beta}_{1b}$ above, why?

- $\beta_{1w} = \beta_{1b}$ and unmeasured confounders
 - within cluster $\Rightarrow \hat{\beta}_{1b}$ unbiased; $\hat{\beta}_{1w}$ biased
 - among clusters \Rightarrow " biased; " unbiased
- $\beta_{1w} \neq \beta_{1b}$

7. what is the difference between a "marginal" or "population-average" MLN and a "conditional" or "cluster-specific" MLN?
When does the distinction matter?

Marginal

$$\log \text{odds}(Y_{ij} = 1) = x_{ij}\beta^M$$

conditional given $b_{\cdot i}$

$$\begin{aligned} \log \text{odds}(Y_{ij} = 1 | b_{\cdot i}) \\ = x_{ij}\beta + z_{ij}b_{\cdot i} \end{aligned}$$

$$|\beta^M| < |\beta^C|$$

attenuation

8. In questions (2) and (3) above, how would you find out whether the sex-specific contextual effects estimates might be biased by the school compositions with respect to variables \bar{z}_{ij} ?

S

g

$$Y_{ij} = \beta_{0i} + \beta_1 (S_{ij} - \bar{S}_{..}) + \beta_{1c} (\bar{S}_i - \bar{S}_{..}) + \beta_2 F_{ij} + \beta_3 (\bar{S}_{i.} - \bar{S}_{..}) F_{ij} + \epsilon_{ij}$$

↓

$$\text{II} + \beta_4 z_{ij}$$

How do β_{1c} , $\beta_2 + \beta_3$, $\text{Var}(\frac{\beta_{0i}}{\beta_{2i}})$ change?

g. As an MPH culminating project, a student conducts a meta-analysis of the following data

STUDY (i)	Effect Estimate ($\hat{\theta}_i$)	$\hat{s}_e(\hat{\theta}_i)$
1	-2	2
2	0	1
3	2	1

Reproduce her forest plot below

FOREST PLOT

study $\hat{\theta}$ $SE_{\hat{\theta}}$

A -2 2

B 0 1

C 2 $\frac{1}{2}$

Pooled

Model: $\hat{\theta}_i = \theta_i + \varepsilon_i$, $\theta_i \stackrel{iid}{\sim} G(0, r^2)$, $\varepsilon_i \sim G(0, SE_{\theta_i}^2)$

Fixed: $r^2 = 0 \Rightarrow \hat{\theta} = \bar{Y}_2 = \left(\frac{1}{2} \right)^2 + \left(\frac{1}{1} \right)^2 + \left(\frac{1}{2} \right)^2 \left[\left(\frac{1}{2} \right)^2 \cdot 2 + 1 \cdot 0 + \left(\frac{1}{2} \right)^2 \cdot 2 \right] = \frac{1}{2} \cdot \left(\frac{1}{4} \right) + \left(\frac{1}{2} \right) = \frac{24}{21} \cdot \frac{15}{21} = \frac{39}{21} = 1.87 = 1.43$

$SE(\hat{\theta}) = \sqrt{\frac{1}{2} \left[\frac{1}{2} + 1 + \left(\frac{1}{2} \right)^2 \right]} = 0.43$

Random:

$$T^2 = 1: \tilde{\theta}_{(1)} = \left[\frac{1}{5} + \frac{1}{2} + \frac{1}{54} \right]^{-1} \left[-\frac{2}{5} + \frac{2}{54} \right]$$

$$= 0.8$$

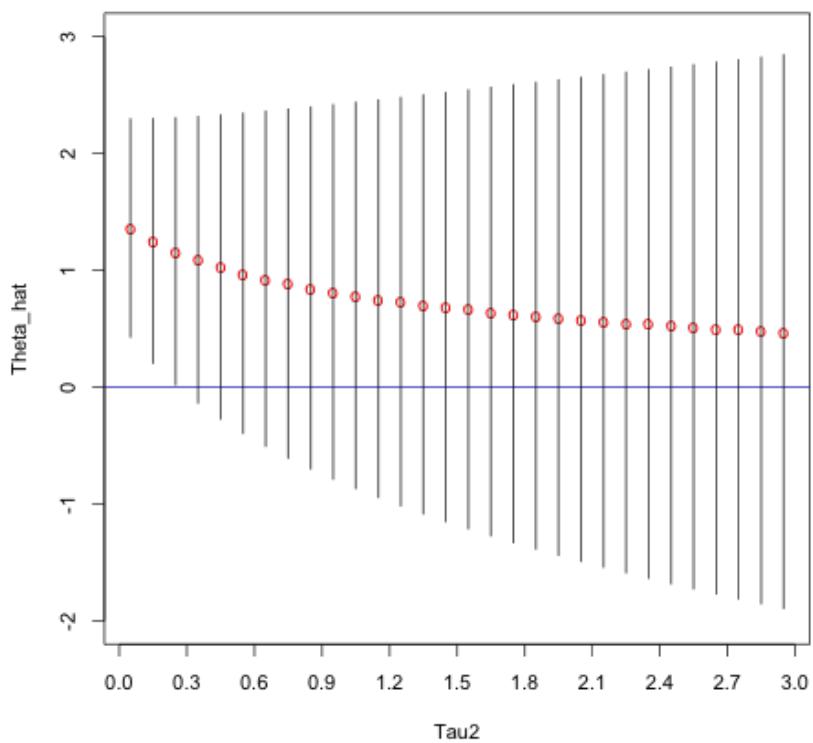
$$SE_{\hat{\theta}_{(1)}} = \left(\frac{1}{5} + \frac{1}{2} + \frac{1}{54} \right)^{1/2} = .82$$

Can't trust estimate of T^2 . Use informative prior for T^2 , be Bayesian!

First, look at $\hat{\theta}_{(g^2)}$ for a range of plausible values of τ^2

$$\text{Estimate } \hat{\tau}^2 = \frac{1}{2} \sum_{j=1}^3 (\hat{\theta}_j - \bar{\hat{\theta}})^2 = \frac{1}{3} \sum_{l=1}^3 (\hat{s}_{\hat{\theta}_l})^2$$
$$= 2.25$$

95% of true θ_i values range from
 $\bar{\hat{\theta}} \pm 4.5$? - scientifically plausible?



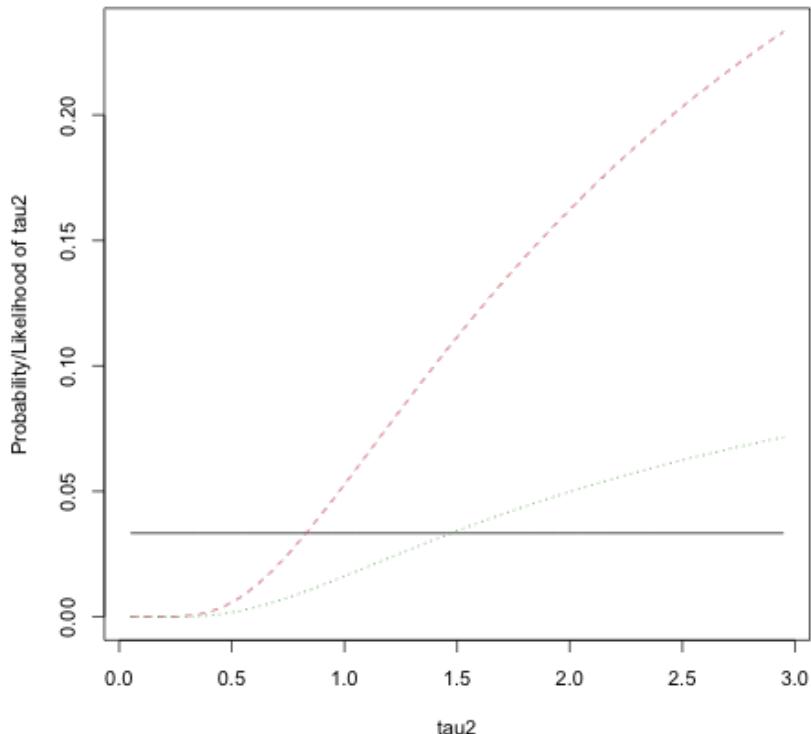
Prior for τ^2 : $U(0, 1)$

Likelihood of $\hat{\tau}^2$ given τ^2

$$2\hat{\tau}/\tau^2 \sim \chi_2^2$$

$$\text{(in general } \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 \text{)}$$

Prior (black), Likelihood (red), Posterior (green) for Tau2



$$\tilde{\gamma}^2 \approx \sum \text{post}(\gamma^2) \cdot \gamma^2$$

$$= 2.13$$

$$\hat{\theta}(2.13) = 0.58$$

-1.51 2.68

10. An HPM student is designing a 2-level study to estimate how the fraction of medical expenditures that are out-of-pocket (OOP) in a year differ between those families in insurance plans A and B. The student has \$105,000. It costs \$5,000 per cluster and \$500 per person in the village to collect the data. She is considering 4 designs:

	n_A villages	n_B subjects/village	n_{total}	$n_{per\ person}$
A	10	5	50	10
B	10	0(10)	10(0)	10
C	15	2	30	4
D	15	0(4)	4(6)	4

How would you advise her as the MLM expert?

Thoughts

- Compare A vs B within and among clusters or only among? (designs B,D)

depends on relative sizes of

$$\text{Var}(B_{0i}), \text{Var}(\varepsilon_{ij})$$

- Model (cartoon)

$$Y_{ij} = \beta_{0i} + \beta_1 \text{Plan}_{ij} + \varepsilon_{ij}$$

$$\beta_{0i} \stackrel{\text{iid}}{\sim} G(\beta_0, \tau^2), \quad \varepsilon_{ij} \stackrel{\text{iid}}{\sim} G(0, \sigma^2)$$

$$\text{Plan}_{ij} = \begin{cases} 1 & A \\ 0 & B \end{cases}$$

Within cluster comparison

$$\text{Var}(\bar{y}_{lA} - \bar{y}_{lB}) = \text{Var}(\beta_0 + \beta_1 + \bar{\epsilon}_{lA} - \bar{\beta}_{0l} - \bar{\epsilon}_{lB}) \\ = \tau^2 + \frac{\sigma^2}{n_a} + \frac{\sigma^2}{n_b}$$

Across cluster comparison

$$\text{Var}(\bar{y}_{lA} - \bar{y}_{l'B}) \quad l \neq l' \\ = 2\tau^2 + \frac{\sigma^2}{n_a} + \frac{\sigma^2}{n_b}$$

In designs A/C :

$$\text{Var}(\hat{\beta}_1) = \frac{\tau^2}{m} + \frac{2\sigma^2}{n}$$

$\sim 10 \text{ or } 15$ ~ 20.5

In designs B/D

$$\text{Var}(\hat{\beta}_1) = \frac{2\tau^2}{m} + \frac{2\sigma^2}{n}$$

$\sim 10 \text{ or } 15$ $\sim 40 \text{ or } 10$