Within Siblings Analysis: Fixed Effects Models

HARVARD
School of Public Health

Department of Epidemiology
Miguel Angel Luque-Fernandez

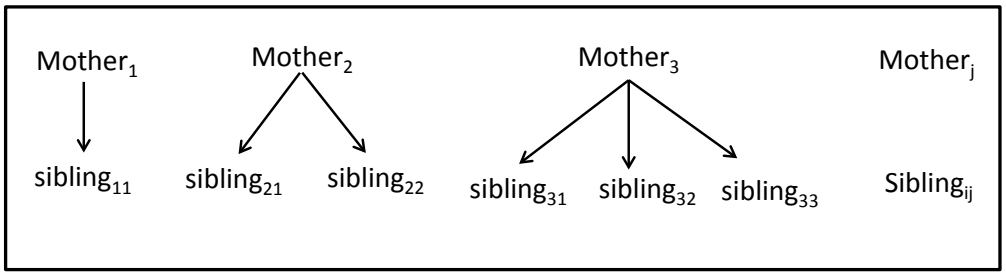
May 2, 2014

Outline

Variance-components models

Introduction

Variance-component models (VCMs) are designed to model and estimate such **within-cluster correlations** in the context of longitudinal and correlated data.

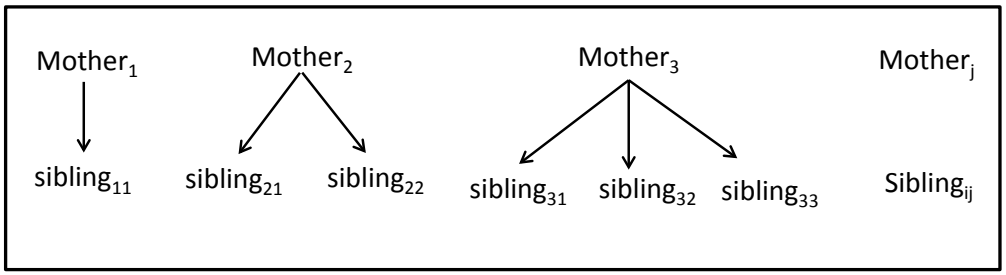


- The effect of mothers can be interpreted as being **random or fixed**.
- In this presentation, we assume that **the effect of mothers is fixed** (i.e., it remains constant across pregnancies).

Variance-components models

Introduction

Variance-component models (VCMs) are designed to model and estimate such **within-cluster correlations** in the context of longitudinal and correlated data.

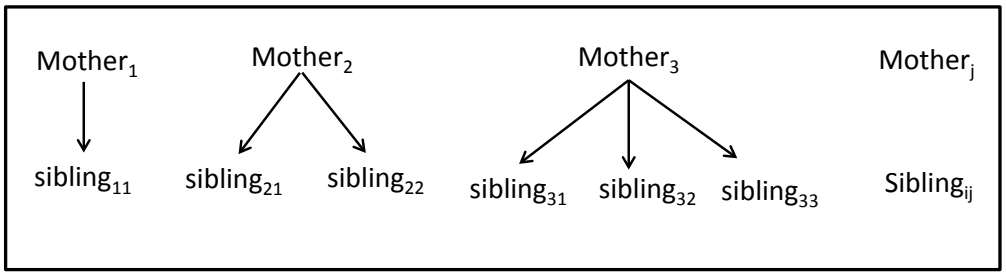


- The effect of mothers can be interpreted as being **random or fixed**.
- In this presentation, we assume that **the effect of mothers is fixed** (i.e., it remains constant across pregnancies).
- Thus, all the genetic and environmental factors associated with mothers remain constant over time and **each mother serves as her own control**.

Variance-components models

Introduction

Variance-component models (VCMs) are designed to model and estimate such **within-cluster correlations** in the context of longitudinal and correlated data.

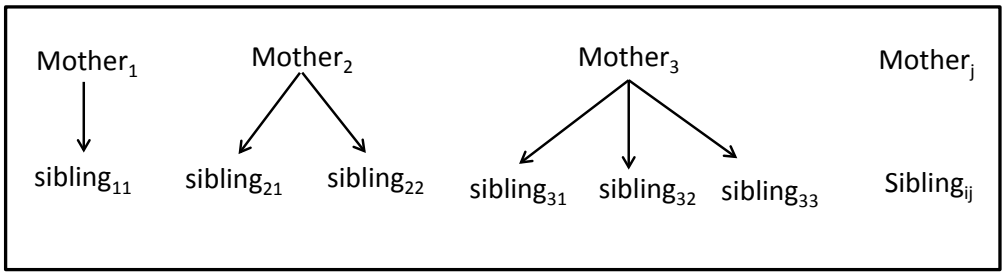


- The effect of mothers can be interpreted as being **random or fixed**.
- In this presentation, we assume that **the effect of mothers is fixed** (i.e., it remains constant across pregnancies).
- Thus, all the genetic and environmental factors associated with mothers remain constant over time and **each mother serves as her own control**.

Variance-components models

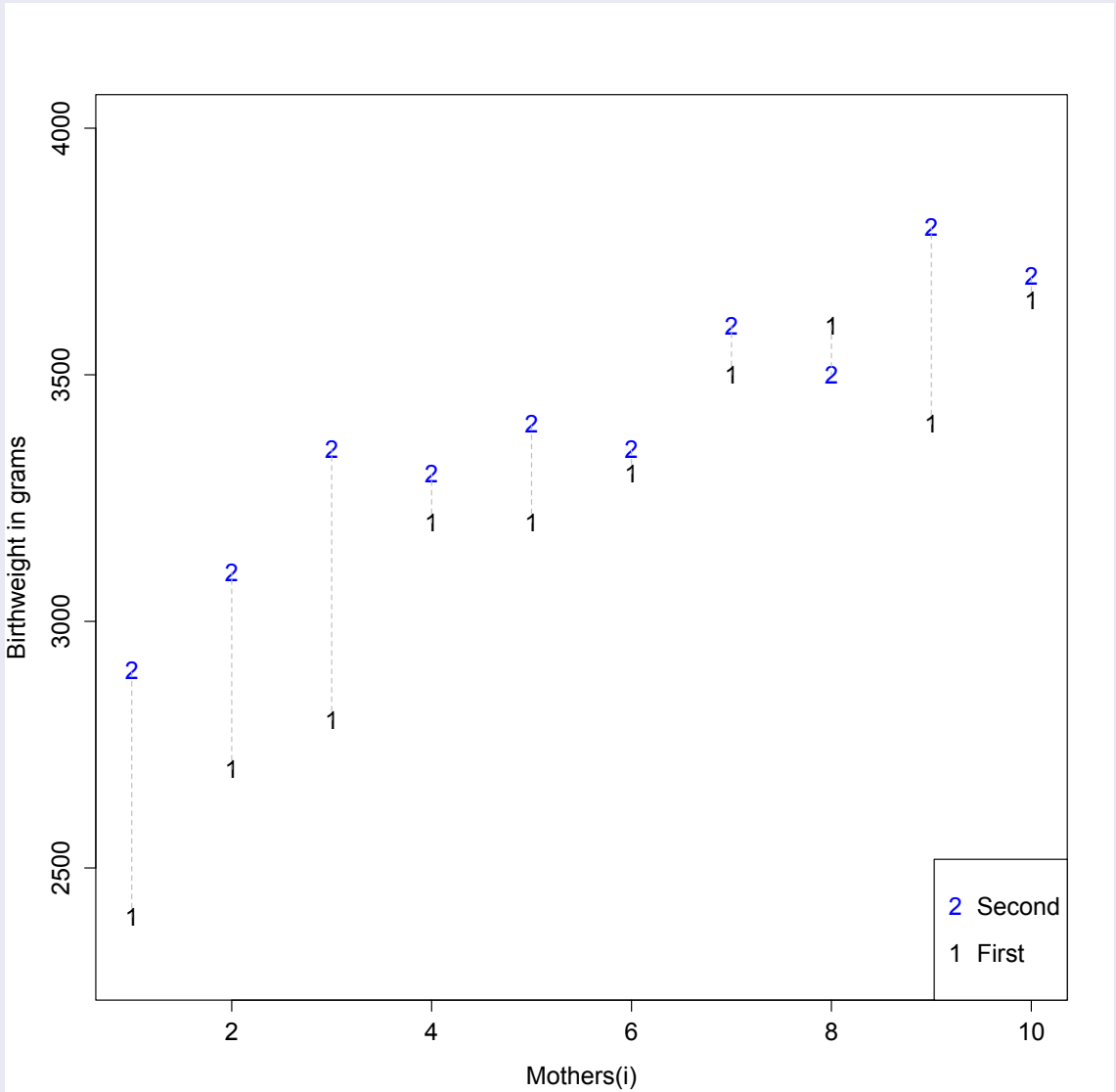
Introduction

Variance-component models (VCMs) are designed to model and estimate such **within-cluster correlations** in the context of longitudinal and correlated data.



- The effect of mothers can be interpreted as being **random or fixed**.
- In this presentation, we assume that **the effect of mothers is fixed** (i.e., it remains constant across pregnancies).
- Thus, all the genetic and environmental factors associated with mothers remain constant over time and **each mother serves as her own control**.

The simplest case: 2k (k = sibling) pairs of repeated measures



The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),

The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,

The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,
- X_{ij}=1 for second infant and 0 for the first, (=j)

The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,
- X_{ij}=1 for second infant and 0 for the first, (=j)
- β: The expected change in birth weight,

The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,
- X_{ij}=1 for second infant and 0 for the first, (=j)
- β: The expected change in birth weight,
- α: The (population) expected response ($\bar{Y}$ birth-weight),

The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,
- X_{ij}=1 for second infant and 0 for the first, (=j)
- β: The expected change in birth weight,
- α: The (population) expected response ($\bar{Y}$ birth-weight),
- U_i: systematic differences between mothers. $U_i \sim N(0, \sigma_u^2)$

The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,
- X_{ij}=1 for second infant and 0 for the first, (=j)
- β: The expected change in birth weight,
- α: The (population) expected response ($\bar{Y}$ birth-weight),
- U_i: systematic differences between mothers. $U_i \sim N(0, \sigma_u^2)$
- e_{ij}: within-siblings error of jth order for the ith mother $e_{ij} \sim N(0, \sigma_e^2)$

The simplest paired design: N Pairs of Siblings

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,
- X_{ij}=1 for second infant and 0 for the first, (=j)
- β: The expected change in birth weight,
- α: The (population) expected response ($\bar{Y}$ birth-weight),
- U_i: systematic differences between mothers. $U_i \sim N(0, \sigma_u^2)$
- e_{ij}: within-siblings error of jth order for the ith mother $e_{ij} \sim N(0, \sigma_e^2)$

The simplest paired design: N Pairs of Siblings

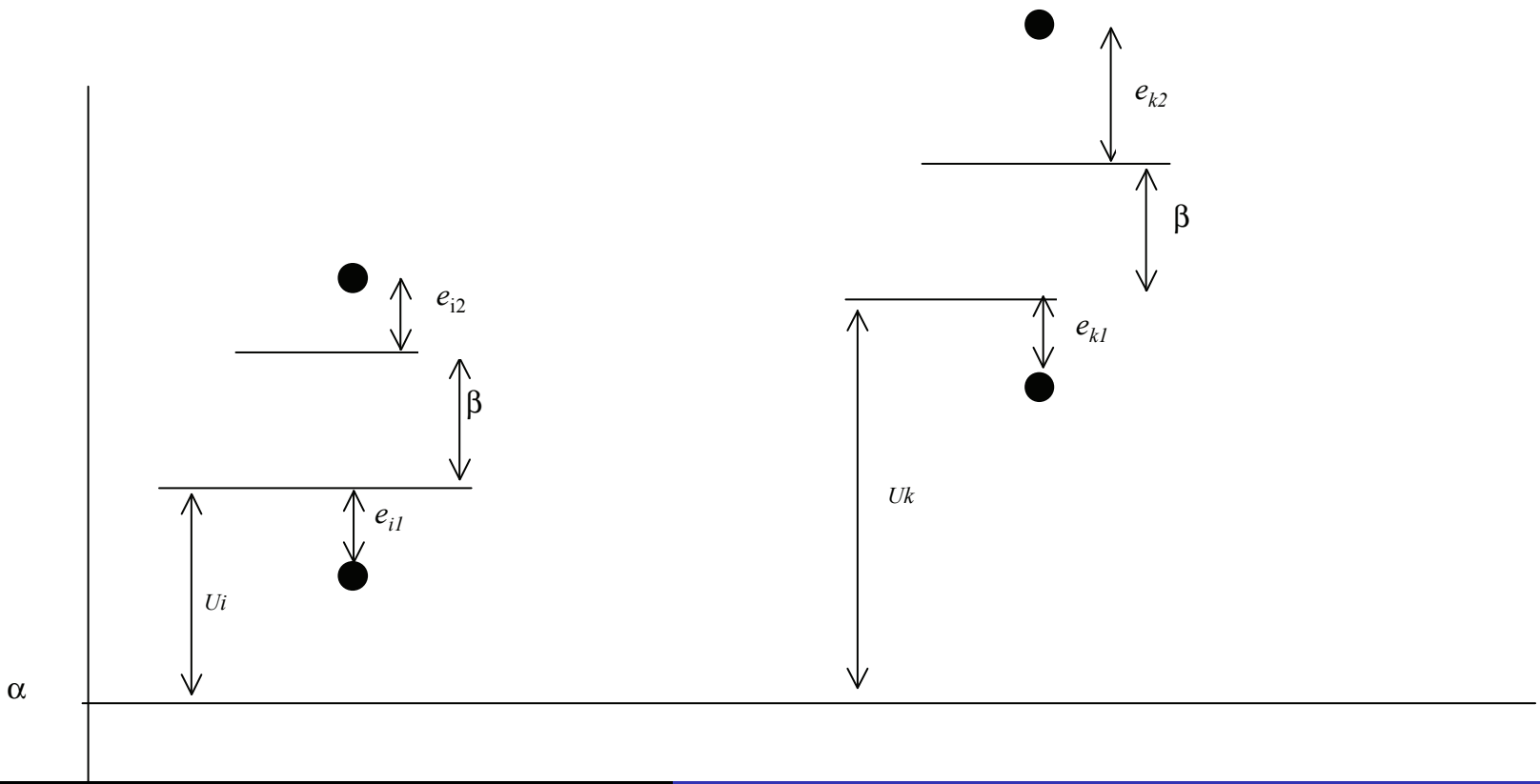
The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Where:

- **Mothers** i=1,...,N
- **Siblings** j=0,1 (0=First and 1=Second),
- Y_{ij}: Birth weight of the infant of mother **i** when delivering the **j** infant,
- X_{ij}=1 for second infant and 0 for the first, (=j)
- β: The expected change in birth weight,
- α: The (population) expected response ($\bar{Y}$ birth-weight),
- U_i: systematic differences between mothers. $U_i \sim N(0, \sigma_u^2)$
- e_{ij}: within-siblings error of jth order for the ith mother $e_{ij} \sim N(0, \sigma_e^2)$

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij} \qquad \text{Cov}(e_{i1}; e_{i0})=0$$



VCMs: Fixed Effect Model

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Within-mothers differences (note that we can also use between mothers dummies variables):

- $Y_{i1} - Y_{i0} = (\alpha + \beta + U_i + e_{i1}) - (\alpha + U_i + e_{i0}) = \beta + (e_{i1} - e_{i0})$
- $E[Y_{i1} - Y_{i0}] = \beta + E[e_{i1} - e_{i0}] = \beta$

VCMs: Fixed Effect Model

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Within-mothers differences (note that we can also use between mothers dummies variables):

- $Y_{i1} - Y_{i0} = (\alpha + \beta + U_i + e_{i1}) - (\alpha + U_i + e_{i0}) = \beta + (e_{i1} - e_{i0})$
- $E[Y_{i1} - Y_{i0}] = \beta + E[e_{i1} - e_{i0}] = \beta$
- $Var(Y_{i1} - Y_{i0}) = Var(e_{i1} - e_{i0}) = 2\sigma_e^2$

VCMs: Fixed Effect Model

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Within-mothers differences (note that we can also use between mothers dummies variables):

- $Y_{i1} - Y_{i0} = (\alpha + \beta + U_i + e_{i1}) - (\alpha + U_i + e_{i0}) = \beta + (e_{i1} - e_{i0})$
- $E[Y_{i1} - Y_{i0}] = \beta + E[e_{i1} - e_{i0}] = \beta$
- $Var(Y_{i1} - Y_{i0}) = Var(e_{i1} - e_{i0}) = 2\sigma_e^2$

Estimates:

- $\hat{\beta} = \sum_{i=1}^{N}(Y_{i1} - Y_{i0})/N = \bar{Y}_1 - \bar{Y}_0$

VCMs: Fixed Effect Model

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Within-mothers differences (note that we can also use between mothers dummies variables):

- $Y_{i1} - Y_{i0} = (\alpha + \beta + U_i + e_{i1}) - (\alpha + U_i + e_{i0}) = \beta + (e_{i1} - e_{i0})$
- $E[Y_{i1} - Y_{i0}] = \beta + E[e_{i1} - e_{i0}] = \beta$
- $Var(Y_{i1} - Y_{i0}) = Var(e_{i1} - e_{i0}) = 2\sigma_e^2$

Estimates:

- $\hat{\beta} = \sum_{i=1}^{N}(Y_{i1} - Y_{i0})/N = \bar{Y}_1 - \bar{Y}_0$
- $Var(\hat{\beta}) = \frac{2\sigma_e^2}{N}$

VCMs: Fixed Effect Model

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Within-mothers differences (note that we can also use between mothers dummies variables):

- $Y_{i1} - Y_{i0} = (\alpha + \beta + U_i + e_{i1}) - (\alpha + U_i + e_{i0}) = \beta + (e_{i1} - e_{i0})$
- $E[Y_{i1} - Y_{i0}] = \beta + E[e_{i1} - e_{i0}] = \beta$
- $Var(Y_{i1} - Y_{i0}) = Var(e_{i1} - e_{i0}) = 2\sigma_e^2$

Estimates:

- $\hat{\beta} = \sum_{i=1}^{N}(Y_{i1} - Y_{i0})/N = \bar{Y}_1 - \bar{Y}_0$
- $Var(\hat{\beta}) = \frac{2\sigma_e^2}{N}$

σ_e^2 only involves within-mothers variation *(FIXED EFFECT MODEL)*

VCMs: Fixed Effect Model

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

Within-mothers differences (note that we can also use between mothers dummies variables):

- $Y_{i1} - Y_{i0} = (\alpha + \beta + U_i + e_{i1}) - (\alpha + U_i + e_{i0}) = \beta + (e_{i1} - e_{i0})$
- $E[Y_{i1} - Y_{i0}] = \beta + E[e_{i1} - e_{i0}] = \beta$
- $Var(Y_{i1} - Y_{i0}) = Var(e_{i1} - e_{i0}) = 2\sigma_e^2$

Estimates:

- $\hat{\beta} = \sum_{i=1}^{N}(Y_{i1} - Y_{i0})/N = \bar{Y}_1 - \bar{Y}_0$
- $Var(\hat{\beta}) = \frac{2\sigma_e^2}{N}$

σ_e^2 only involves within-mothers variation *(FIXED EFFECT MODEL)*

Within-sibling correlations: Intra cluster correlation (ICC)

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

- Within-siblings correlation (RHO, ρ): "Intra Cluster Correlation"
- $\rho = Corr(Y_{i1}, Y_{i0}) = Corr(U_i + e_{i1}, U_i + e_{i0})$

Within-sibling correlations: Intra cluster correlation (ICC)

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

- Within-siblings correlation (RHO, ρ): "Intra Cluster Correlation"
- $\rho = Corr(Y_{i1}, Y_{i0}) = Corr(U_i + e_{i1}, U_i + e_{i0})$

ICC

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \quad ; \text{Therefore} \quad \sigma_e^2 = \sigma_T^2(1-\rho)$$

Within-sibling correlations: Intra cluster correlation (ICC)

The Model

$$Y_{ij} = \alpha + \beta X_{ij} + U_i + e_{ij}$$

- Within-siblings correlation (RHO, ρ): "Intra Cluster Correlation"
- $\rho = Corr(Y_{i1}, Y_{i0}) = Corr(U_i + e_{i1}, U_i + e_{i0})$

ICC

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \quad ; \text{Therefore} \quad \sigma_e^2 = \sigma_T^2(1-\rho)$$

ICC as link between paired and independent analysis

Take message

Within-cluster (mothers) correlation increases precision for estimating within-siblings effects. So, we can obtain greater benefit pairing if ICC (ρ) is larger.

The proof

Paired: $\frac{2\sigma_e^2}{N} = \frac{2\sigma_T^2(1-\rho)}{N}$ Independent: $\frac{2(\sigma_u^2+\sigma_e^2)}{N} = \frac{2\sigma_T^2}{N}$

ICC as link between paired and independent analysis

Take message

Within-cluster (mothers) correlation increases precision for estimating within-siblings effects. So, we can obtain greater benefit pairing if ICC (ρ) is larger.

The proof

Paired: $\frac{2\sigma_e^2}{N} = \frac{2\sigma_T^2(1-\rho)}{N}$ Independent: $\frac{2(\sigma_u^2 + \sigma_e^2)}{N} = \frac{2\sigma_T^2}{N}$

Analytical methods for more than 2k siblings

Mean-centered transformation 'Demeaning'

$$Y_{ij}^{*} = \beta X_{ij}^{'*} + e_{ij}^{*},$$

Where

Mean-centered outcome and covariates

$$Y_{ij}^{*} = Y_{ij} - \bar{Y}_i\,; \qquad \bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

$$X_{ijk}^{*} = X_{ijk} - \bar{X}_{ik}\,; \qquad \bar{X}_{ik} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ijk}$$

Analytical methods for more than 2k siblings

Mean-centered transformation 'Demeaning'

$$Y_{ij}^{*} = \beta X_{ij}^{'*} + e_{ij}^{*},$$

Where

Mean-centered outcome and covariates

$$Y_{ij}^{*} = Y_{ij} - \bar{Y}_i\,; \qquad \bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

$$X_{ijk}^{*} = X_{ijk} - \bar{X}_{ik}\,; \qquad \bar{X}_{ik} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ijk}$$

Fixed Effect Models summary (FEMs)

FEMs summary

- Account for all between mothers variation by **holding the effect of mothers constant** (fixed intercept in the model).
- All inference is therefore **within-siblings**.

Fixed Effect Models summary (FEMs)

FEMs summary

- Account for all between mothers variation by **holding the effect of mothers constant** (fixed intercept in the model).
- All inference is therefore **within-siblings**.
- Can be implemented either by **dummy variables** for n-1 mothers or **subtracting mothers mean** birth weights from all the siblings of the same mother.

Fixed Effect Models summary (FEMs)

FEMs summary

- Account for all between mothers variation by **holding the effect of mothers constant** (fixed intercept in the model).
- All inference is therefore **within-siblings**.
- Can be implemented either by **dummy variables** for n-1 mothers or **subtracting mothers mean** birth weights from all the siblings of the same mother.

Fixed Effect Models summary (FEMs)

FEMs summary

- Account for all between mothers variation by **holding the effect of mothers constant** (fixed intercept in the model).
- All inference is therefore **within-siblings**.
- Can be implemented either by **dummy variables** for n-1 mothers or **subtracting mothers mean** birth weights from all the siblings of the same mother.

FEMs Merits

Merits

- Control for unmeasured time-invariant characteristics of individuals (genetic and environmental factors),
- Each subject (mothers) truly serves as its own control,

FEMs Merits

Merits

- Control for unmeasured time-invariant characteristics of individuals (genetic and environmental factors),
- Each subject (mothers) truly serves as its own control,
- Assuming that the effects on the response of time-invariant characteristics remain constant over time, FE models provide unbiased estimates of time-varying covariates.

FEMs Merits

Merits

- Control for unmeasured time-invariant characteristics of individuals (genetic and environmental factors),
- Each subject (mothers) truly serves as its own control,
- Assuming that the effects on the response of time-invariant characteristics remain constant over time, FE models provide unbiased estimates of time-varying covariates.

FEMs Merits

Merits

- Control for unmeasured time-invariant characteristics of individuals (genetic and environmental factors),
- Each subject (mothers) truly serves as its own control,
- Assuming that the effects on the response of time-invariant characteristics remain constant over time, FE models provide unbiased estimates of time-varying covariates.

Fixed Effects Models Limitations

FEMs Limitations

- FE cannot estimate the effects of time-invariant covariates,
- The effect of time-invariant covariates is assumed to remain constant over time,

Fixed Effects Models Limitations

FEMs Limitations

- FE cannot estimate the effects of time-invariant covariates,
- The effect of time-invariant covariates is assumed to remain constant over time,
- Need of strict exogeneity, which prohibits some types of feedback from past outcomes to current covariates and current outcome to future covariates (reverse causality -smoking and prematurity),

Fixed Effects Models Limitations

FEMs Limitations

- FE cannot estimate the effects of time-invariant covariates,
- The effect of time-invariant covariates is assumed to remain constant over time,
- Need of strict exogeneity, which prohibits some types of feedback from past outcomes to current covariates and current outcome to future covariates (reverse causality -smoking and prematurity),
- Trade-off between bias and efficiency. Lack of precision particularly if there is little within-cluster variation,

Fixed Effects Models Limitations

FEMs Limitations

- FE cannot estimate the effects of time-invariant covariates,
- The effect of time-invariant covariates is assumed to remain constant over time,
- Need of strict exogeneity, which prohibits some types of feedback from past outcomes to current covariates and current outcome to future covariates (reverse causality -smoking and prematurity),
- Trade-off between bias and efficiency. Lack of precision particularly if there is little within-cluster variation,
- FE models yields invalid inference in presence of missing at random (MAR) data.

Fixed Effects Models Limitations

FEMs Limitations

- FE cannot estimate the effects of time-invariant covariates,
- The effect of time-invariant covariates is assumed to remain constant over time,
- Need of strict exogeneity, which prohibits some types of feedback from past outcomes to current covariates and current outcome to future covariates (reverse causality -smoking and prematurity),
- Trade-off between bias and efficiency. Lack of precision particularly if there is little within-cluster variation,
- FE models yields invalid inference in presence of missing at random (MAR) data.

Fixed Effects Models Limitations

FEMs Limitations

- FE cannot estimate the effects of time-invariant covariates,
- The effect of time-invariant covariates is assumed to remain constant over time,
- Need of strict exogeneity, which prohibits some types of feedback from past outcomes to current covariates and current outcome to future covariates (reverse causality -smoking and prematurity),
- Trade-off between bias and efficiency. Lack of precision particularly if there is little within-cluster variation,
- FE models yields invalid inference in presence of missing at random (MAR) data.

Thank You

mluquefe@hsph.harvard.edu

Alhambra: Granada, Spain