Including family effects in multilevel models for pupil progress

Rebecca Pillinger, Jon Rasbash, George Leckie and Jenny Jenkins

1. Partitioning variation in progress

What do we already know?

	①		②	③	④		⑤		⑥	⑦		⑧	⑨
Response	Prog	Prog	Prog	Att	Att	Prog	Att	Prog	Att	Att	Prog	Cog	Cog
Predictors	Y	Y	Y	N	N	Y	N	Y	N	N	Y	N	N
LEA			0.2	3									
Neighbhd				2	20	4	14	1	5	4	1		0.20
Secondary	5—20		5	22		fixed	7	1	20	23	4		
Primary		5—20								3	7		
Cohort			3										0.21
Pupil	80—95	80—95	93	73	80	96	79	98	75	70	88		
MZ twins												0.86	0.78
DZ twins												0.60	0.64
Full sibs												0.47	0.51

① Typical school effectiveness studies, e.g. Goldstein et al. (2007) (primary schools) and Leckie (2008) (secondary schools)

② Yang & Woodhouse (2001), progress from GCSE to A-level

③ Fielding et al. (2006)

④ Garner & Raudenbush (1991); predictors include family background, neighbourhood social deprivation and school fixed effects

⑤ Raudenbush (1993); reanalysis of Garner & Raudenbush (1991)

⑥ Leckie (2008)

⑦ Leckie (2008)

⑧ Bouchard & McGue (1981); metaanalysis of 110 studies

⑨ Duncan et al. (2001); US data; response is Peabody Picture Vocabulary Test

Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Developmental Psychology

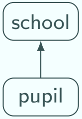
- Models usually have children within families
- Researchers recognise that really these models partition into shared environment and non-shared environment
- What is the shared environment?
- In other words, How much of the shared environment is family, school and area?

Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram



Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

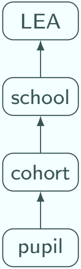
Classification diagram

Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram

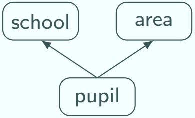


Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram

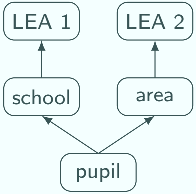


Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram

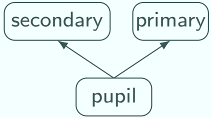


Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram

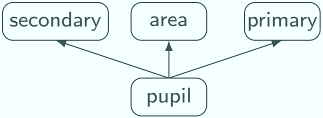


Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram

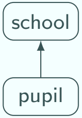


Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram



Model

$$y_{ij} = \alpha + \beta x_{ij} + u_j + e_{ij},$$

$$u_j \sim \mathrm{N}\left(0, \sigma_u^2\right), \qquad i = 1, \ldots, n_j$$
$$e_{ij} \sim \mathrm{N}\left(0, \sigma_e^2\right), \qquad j = 1, \ldots, J$$

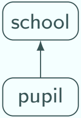
(B)

Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram



Model

$$\mathbf{GCSE}_{ij} = \alpha + \beta\mathbf{pretest}_{ij} + u_j + e_{ij},$$

$$u_j \sim \mathrm{N}\left(0, \sigma_u^2\right), \qquad i = 1, \ldots, n_j$$
$$e_{ij} \sim \mathrm{N}\left(0, \sigma_e^2\right), \qquad j = 1, \ldots, J$$

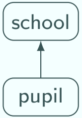
(B)

Previous studies

School effectiveness

- Models usually have pupils within schools (2 levels)
- There have been studies that also included area or primary school
- But no studies have included family
- The largest component of variation in these models is the pupil level
- How much of that is really family level?

Classification diagram



Model

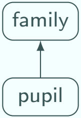
$$\mathbf{GCSE}_{ij} = \alpha + \beta_1 \mathbf{pretest}_{ij} + \beta_2 x_{ij} + u_j + e_{ij},$$

$$u_j \sim \mathrm{N}\left(0, \sigma_u^2\right), \qquad i = 1, \ldots, n_j$$
$$e_{ij} \sim \mathrm{N}\left(0, \sigma_e^2\right), \qquad j = 1, \ldots, J$$

(B)

Previous studies

Classification diagram

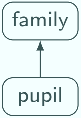


Developmental Psychology

- Models usually have children within families
- Researchers recognise that really these models partition into shared environment and non-shared environment
- What is the shared environment?
- In other words, How much of the shared environment is family, school and area?

Previous studies

Classification diagram



Model

$$y_{ij} = \alpha + u_j + e_{ij} + g_{ij}$$

$$u_j \sim N(0, \sigma_u^2)$$
$$e_{ij} \sim N(0, \sigma_e^2)$$
$$g_{ij} \sim N(0, \sigma_g^2)$$

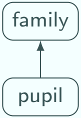
$$\text{Cov}(g_{i_1 j}, g_{i_2 j}) = r_{(i_1 j, i_2 j)} \sigma_g^2$$

Developmental Psychology

- Models usually have children within families
- Researchers recognise that really these models partition into shared environment and non-shared environment
- What is the shared environment?
- In other words, How much of the shared environment is family, school and area?

Previous studies

Classification diagram



Model

$$y_{ij} = \alpha + d_j u_{1j} + d_j e_{1ij} + (1 - d_j) e_{2ij}$$

$$u_{1j} \sim \mathrm{N}\left(0, \sigma_u^2\right),$$

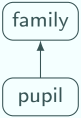
$$\begin{bmatrix} e_{1ij} \\ e_{2ij} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{e1}^2 & \\ 0 & \sigma_{e2}^2 \end{bmatrix} \right) \qquad \text{(A)}$$

Developmental Psychology

- Models usually have children within families
- Researchers recognise that really these models partition into shared environment and non-shared environment
- What is the shared environment?
- In other words, How much of the shared environment is family, school and area?

Previous studies

Classification diagram



Model

$$\textbf{GCSE}_{ij} = \alpha + \textbf{twin}_j u_{1j} + \textbf{twin}_j e_{1ij} + \textbf{nontwin}_j e_{2ij}$$

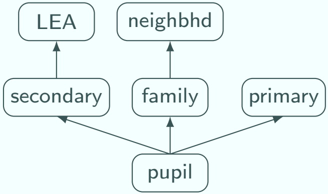
$$u_{1j} \sim \mathrm{N}\left(0, \sigma^2_u\right),$$
$$\begin{bmatrix} e_{1ij} \\ e_{2ij} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{e1} & \\ 0 & \sigma^2_{e2} \end{bmatrix} \right) \quad \text{(A)}$$

Developmental Psychology

- Models usually have children within families
- Researchers recognise that really these models partition into shared environment and non-shared environment
- What is the shared environment?
- In other words, How much of the shared environment is family, school and area?

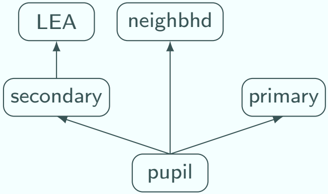
Our model

Classification diagram



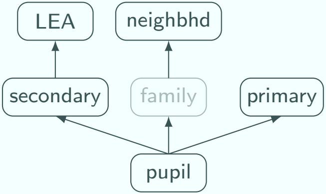
Our model

Classification diagram



Our model

Classification diagram

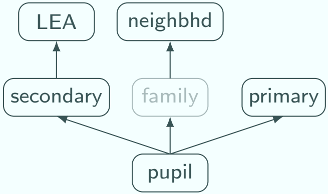


Our model

Model

$$y_i = \alpha + \beta x_i + u^{(6)}_{\text{LEA}(i)} + u^{(5)}_{\text{sec}(i)} + u^{(4)}_{\text{nbhd}(i)} + u^{(3)}_{\text{pri}(i)} + d_{\text{fam}(i)} u^{(2)}_{\text{fam}(i)} + d_{\text{fam}(i)} e_{1i} + \left(1 - d_{\text{fam}(i)}\right) e_{2i}$$

Classification diagram



$$u^{(6)}_{\text{LEA}(i)} \sim \text{N}\left(0, \sigma^2_{u(6)}\right)$$

$$\vdots \qquad \vdots$$

$$u^{(3)}_{\text{pri}(i)} \sim \text{N}\left(0, \sigma^2_{u(3)}\right)$$

$$u^{(2)}_{\text{fam}(i)} \sim \text{N}\left(0, \sigma^2_{u(2)}\right)$$

$$\begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} \sim \text{N}\left(0, \begin{bmatrix} \sigma^2_{e1} & \\ 0 & \sigma^2_{e2} \end{bmatrix}\right)$$

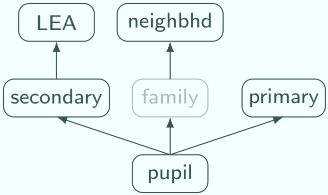
(C)

Our model

Model

$$\mathbf{GCSE}_i = \alpha + \beta_1 \mathbf{pretest}_i + \beta_2 \mathbf{twin}_i + \beta_3 \mathbf{pretest} \cdot \mathbf{twin}_i + u^{(6)}_{\text{LEA}(i)} + u^{(5)}_{\text{sec}(i)} + u^{(4)}_{\text{nbhd}(i)} + u^{(3)}_{\text{pri}(i)} + \mathbf{twin}_{\text{fam}(i)} u^{(2)}_{\text{fam}(i)} + \mathbf{twin}_{\text{fam}(i)} e_{1i} + \mathbf{nontwin}_{\text{fam}(i)} e_{2i}$$

Classification diagram



$$u^{(6)}_{\text{LEA}(i)} \sim \text{N}\left(0, \sigma^2_{u(6)}\right)$$

$$\vdots \qquad \vdots$$

$$u^{(3)}_{\text{pri}(i)} \sim \text{N}\left(0, \sigma^2_{u(3)}\right)$$

$$u^{(2)}_{\text{fam}(i)} \sim \text{N}\left(0, \sigma^2_{u(2)}\right)$$

$$\begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} \sim \text{N}\left(0, \begin{bmatrix} \sigma^2_{e1} & \\ 0 & \sigma^2_{e2} \end{bmatrix}\right)$$

(C)

Our data

Sample

All pupils in

- England
- state schools
- 2007 GCSE cohort

Our data

Sample

All pupils in

- England
- state schools
- 2007 GCSE cohort

Variables

- Test scores from the NPD
  - GCSE (our response) and
  - key stage 2 (KS2)
- Background characteristics from PLASC
  - age
  - gender
  - ethnicity
  - FSM eligibility
  - SEN
  - EAL
- ONS data on LSOAs
  - IDACI

All continuous variables have been standardized

Our data

Sample

All pupils in

- England
- state schools
- 2007 GCSE cohort

Levels

- The data records which
  - LEA
  - secondary school
  - primary school
  - area (LSOA)

  each pupil belongs to
- But not which family

Variables

- Test scores from the NPD
  - GCSE (our response) and
  - key stage 2 (KS2)
- Background characteristics from PLASC
  - age
  - gender
  - ethnicity
  - FSM eligibility
  - SEN
  - EAL
- ONS data on LSOAs
  - IDACI

All continuous variables have been standardized

Our data

Identifying twins

- We get the family level by identifying twin pairs
- by matching on time invariant characteristics
  - date of birth
  - ethnicity
  - EAL

  and pattern of time-varying characteristics
  - postcode sector
  - FSM eligiblity

How successful is this?

- 11.54 twin births per 1000 maternities in 1990 & 1991
- 9.37 twin pairs per 1000 families in our matching
- We may also have labelled some unrelated pupils as a ‘twin pair’
- Calculation suggests around 10% of ‘twin pairs’ will be coincidental matches

Size of dataset

551,220 pupils	30507 LSOAs	3099 secondaries
5116 twin pairs	14765 primaries	149 LEAs

Results

	Model A		Model B		Model C		Model D	
cons	−0.003	(0.001)	−0.003	(0.001)	0.001	(0.008)	−0.039	(0.007)
twin	0.177	(0.008)	0.179	(0.007)	0.162	(0.007)	0.154	(0.007)
pretest	0.730	(0.001)	0.729	(0.001)	0.701	(0.001)	0.641	(0.001)
pretest.twin	−0.040	(0.007)	0.000	(0.007)	−0.027	(0.006)	−0.020	(0.006)
female							0.184	(0.002)
Asian							0.429	(0.005)
Black							0.225	(0.006)
Chinese							0.556	(0.015)
Mixed							0.045	(0.005)
Other							0.403	(0.010)
FSM							−0.248	(0.003)
age							−0.012	(0.000)
SEN							−0.231	(0.003)
IDACI							−0.103	(0.001)
LEA					0.005	(0.001)	0.005	(0.001)
Secondary			0.065	(0.002)	0.043	(0.001)	0.035	(0.001)
Primary					0.035	(0.001)	0.025	(0.000)
LSOA					0.008	(0.000)	0.002	(0.000)
Family (twin)	0.238	(0.007)			0.168	(0.005)	0.157	(0.005)
Pupil (twin)	0.160	(0.003)			0.157	(0.003)	0.150	(0.003)
Pupil (non-twin)	0.468	(0.001)	0.402	(0.002)	0.383	(0.001)	0.357	(0.001)

Using MCMC; 450,500 iterations and a burn-in of 50,000

Results

	Model A		Model B		Model C		Model D	
cons	−0.003	(0.001)	−0.003	(0.001)	0.001	(0.008)	−0.039	(0.007)
twin	0.177	(0.008)	0.179	(0.007)	0.162	(0.007)	0.154	(0.007)
pretest	0.730	(0.001)	0.729	(0.001)	0.701	(0.001)	0.641	(0.001)
pretest.twin	−0.040	(0.007)	0.000	(0.007)	−0.027	(0.006)	−0.020	(0.006)
female							0.184	(0.002)
Asian							0.429	(0.005)
Black							0.225	(0.006)
Chinese							0.556	(0.015)
Mixed							0.045	(0.005)
Other							0.403	(0.010)
FSM							−0.248	(0.003)
age							−0.012	(0.000)
SEN							−0.231	(0.003)
IDACI							−0.103	(0.001)
LEA					0.005	(0.001)	0.005	(0.001)
Secondary			0.065	(0.002)	0.043	(0.001)	0.035	(0.001)
Primary					0.035	(0.001)	0.025	(0.000)
LSOA					0.008	(0.000)	0.002	(0.000)
Family (twin)	0.238	(0.007)			0.168	(0.005)	0.157	(0.005)
Pupil (twin)	0.160	(0.003)			0.157	(0.003)	0.150	(0.003)
Pupil (non-twin)	0.468	(0.001)	0.402	(0.002)	0.383	(0.001)	0.357	(0.001)

Using MCMC; 450,500 iterations and a burn-in of 50,000

Results

	Model A		Model B		Model C		Model D	
cons	−0.003	(0.001)	−0.003	(0.001)	0.001	(0.008)	−0.039	(0.007)
twin	0.177	(0.008)	0.179	(0.007)	0.162	(0.007)	0.154	(0.007)
pretest	0.730	(0.001)	0.729	(0.001)	0.701	(0.001)	0.641	(0.001)
pretest.twin	−0.040	(0.007)	0.000	(0.007)	−0.027	(0.006)	−0.020	(0.006)
female							0.184	(0.002)
Asian							0.429	(0.005)
Black							0.225	(0.006)
Chinese							0.556	(0.015)
Mixed							0.045	(0.005)
Other							0.403	(0.010)
FSM							−0.248	(0.003)
age							−0.012	(0.000)
SEN							−0.231	(0.003)
IDACI							−0.103	(0.001)
LEA					0.005	(0.001)	0.005	(0.001)
Secondary			0.065	(0.002)	0.043	(0.001)	0.035	(0.001)
Primary					0.035	(0.001)	0.025	(0.000)
LSOA					0.008	(0.000)	0.002	(0.000)
Family (twin)	0.238	(0.007)			0.168	(0.005)	0.157	(0.005)
Pupil (twin)	0.160	(0.003)			0.157	(0.003)	0.150	(0.003)
Pupil (non-twin)	0.468	(0.001)	0.402	(0.002)	0.383	(0.001)	0.357	(0.001)

Using MCMC; 450,500 iterations and a burn-in of 50,000

Results

	Model A		Model B		Model C		Model D	
cons	−0.003	(0.001)	−0.003	(0.001)	0.001	(0.008)	−0.039	(0.007)
twin	0.177	(0.008)	0.179	(0.007)	0.162	(0.007)	0.154	(0.007)
pretest	0.730	(0.001)	0.729	(0.001)	0.701	(0.001)	0.641	(0.001)
pretest.twin	−0.040	(0.007)	0.000	(0.007)	−0.027	(0.006)	−0.020	(0.006)
female							0.184	(0.002)
Asian							0.429	(0.005)
Black							0.225	(0.006)
Chinese							0.556	(0.015)
Mixed							0.045	(0.005)
Other							0.403	(0.010)
FSM							−0.248	(0.003)
age							−0.012	(0.000)
SEN							−0.231	(0.003)
IDACI							−0.103	(0.001)
LEA					0.005	(0.001)	0.005	(0.001)
Secondary			0.065	(0.002)	0.043	(0.001)	0.035	(0.001)
Primary					0.035	(0.001)	0.025	(0.000)
LSOA					0.008	(0.000)	0.002	(0.000)
Family (twin)	0.238	(0.007)			0.168	(0.005)	0.157	(0.005)
Pupil (twin)	0.160	(0.003)			0.157	(0.003)	0.150	(0.003)
Pupil (non-twin)	0.468	(0.001)	0.402	(0.002)	0.383	(0.001)	0.357	(0.001)

Using MCMC; 450,500 iterations and a burn-in of 50,000

Variance partitioning coefficients

	A	B	C		D	
			Twins	Non-twins	Twins	Non-twins
LEA			1.2%	1.1%	1.3%	1.2%
Secondary		13.9%	10.3%	9.1%	9.4%	8.3%
Primary			8.4%	7.4%	6.7%	5.9%
LSOA			1.9%	1.7%	0.5%	0.5%
Family	59.8%		40.4%		42.0%	
Pupil	40.2%	86.1%	37.7%	80.8%	40.1%	84.2%

Research questions

Variance partitioning coefficients

	A	B	C		D	
			Twins	Non-twins	Twins	Non-twins
LEA			1.2%	1.1%	1.3%	1.2%
Secondary		13.9%	10.3%	9.1%	9.4%	8.3%
Primary			8.4%	7.4%	6.7%	5.9%
LSOA			1.9%	1.7%	0.5%	0.5%
Family	59.8%		40.4%		42.0%	
Pupil	40.2%	86.1%	37.7%	80.8%	40.1%	84.2%

Research questions

1. How much of the shared environmental variation is due to family, school and area?

Variance partitioning coefficients

	A	B	C		D	
			Twins	Non-twins	Twins	Non-twins
LEA			1.2%	1.1%	1.3%	1.2%
Secondary		13.9%	10.3%	9.1%	9.4%	8.3%
Primary			8.4%	7.4%	6.7%	5.9%
LSOA			1.9%	1.7%	0.5%	0.5%
Family	59.8%		40.4%		42.0%	
Pupil	40.2%	86.1%	37.7%	80.8%	40.1%	84.2%

Research questions

1. How much of the shared environmental variation is due to family, school and area?
2. How much of the 'pupil' level variation in school effectiveness studies is really family level?

Variance partitioning coefficients

	A	B	C		D	
			Twins	Non-twins	Twins	Non-twins
LEA			1.2%	1.1%	1.3%	1.2%
Secondary		13.9%	10.3%	9.1%	9.4%	8.3%
Primary			8.4%	7.4%	6.7%	5.9%
LSOA			1.9%	1.7%	0.5%	0.5%
Family	59.8%		40.4%		42.0%	
Pupil	40.2%	86.1%	37.7%	80.8%	40.1%	84.2%

Research questions

1. How much of the shared environmental variation is due to family, school and area?
2. How much of the 'pupil' level variation in school effectiveness studies is really family level?

Variance partitioning coefficients

	A	B	C		D	
			Twins	Non-twins	Twins	Non-twins
LEA			1.2%	1.1%	1.3%	1.2%
Secondary		13.9%	10.3%	9.1%	9.4%	8.3%
Primary			8.4%	7.4%	6.7%	5.9%
LSOA			1.9%	1.7%	0.5%	0.5%
Family	59.8%		40.4%		42.0%	
Pupil	40.2%	86.1%	37.7%	80.8%	40.1%	84.2%

Research questions

1. How much of the shared environmental variation is due to family, school and area?
2. How much of the 'pupil' level variation in school effectiveness studies is really family level?

What happens when we try to explain some of the variation using pupil, family and LSOA level covariates?

Results

	Model A		Model B		Model C		Model D	
cons	−0.003	(0.001)	−0.003	(0.001)	0.001	(0.008)	−0.039	(0.007)
twin	0.177	(0.008)	0.179	(0.007)	0.162	(0.007)	0.154	(0.007)
pretest	0.730	(0.001)	0.729	(0.001)	0.701	(0.001)	0.641	(0.001)
pretest.twin	−0.040	(0.007)	0.000	(0.007)	−0.027	(0.006)	−0.020	(0.006)
female							0.184	(0.002)
Asian							0.429	(0.005)
Black							0.225	(0.006)
Chinese							0.556	(0.015)
Mixed							0.045	(0.005)
Other							0.403	(0.010)
FSM							−0.248	(0.003)
age							−0.012	(0.000)
SEN							−0.231	(0.003)
IDACI							−0.103	(0.001)
LEA					0.005	(0.001)	0.005	(0.001)
Secondary			0.065	(0.002)	0.043	(0.001)	0.035	(0.001)
Primary					0.035	(0.001)	0.025	(0.000)
LSOA					0.008	(0.000)	0.002	(0.000)
Family (twin)	0.238	(0.007)			0.168	(0.005)	0.157	(0.005)
Pupil (twin)	0.160	(0.003)			0.157	(0.003)	0.150	(0.003)
Pupil (non-twin)	0.468	(0.001)	0.402	(0.002)	0.383	(0.001)	0.357	(0.001)

Using MCMC; 450,500 iterations and a burn-in of 50,000

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Caveats

- Our family effects are purely derived from twin pairs

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Caveats

- Our family effects are purely derived from twin pairs
- The twins are a mix of MZ and DZ so we are not estimating $\sigma_u^2 + \sigma_g^2$

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Caveats

- Our family effects are purely derived from twin pairs
- The twins are a mix of MZ and DZ so we are not estimating $\sigma_u^2 + \sigma_g^2$
- Twins may also be different to full sibling pairs

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Caveats

- Our family effects are purely derived from twin pairs
- The twins are a mix of MZ and DZ so we are not estimating $\sigma_u^2 + \sigma_g^2$
- Twins may also be different to full sibling pairs
  - shared environment in womb

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Caveats

- Our family effects are purely derived from twin pairs
- The twins are a mix of MZ and DZ so we are not estimating $\sigma_u^2 + \sigma_g^2$
- Twins may also be different to full sibling pairs
  - shared environment in womb
  - they may elicit more similar environments

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Caveats

- Our family effects are purely derived from twin pairs
- The twins are a mix of MZ and DZ so we are not estimating $\sigma_u^2 + \sigma_g^2$
- Twins may also be different to full sibling pairs
  - shared environment in womb
  - they may elicit more similar environments
  - have same age sibling

Interpretation

Summary

- Around a third of the family level variation in Model A is really school or area level (mostly school)
- Around half the pupil level variation in Model B is really family level
- The covariates explain some variation at most levels
- Family and pupil still make up the largest, roughly equal proportions of variation
- Both school levels also remain important

Caveats

- Our family effects are purely derived from twin pairs
- The twins are a mix of MZ and DZ so we are not estimating $\sigma_u^2 + \sigma_g^2$
- Twins may also be different to full sibling pairs
  - shared environment in womb
  - they may elicit more similar environments
  - have same age sibling
- To what extent can we generalise to non-nuclear families?

2. What happens under stress?

Variance functions for stress

Data

- Data is from previous cohort, who took GCSEs in 2006
- Postcodes with more than 2 students excluded
- Continuous variables not standardized

Model

Variance functions for stress

Data

- Data is from previous cohort, who took GCSEs in 2006
- Postcodes with more than 2 students excluded
- Continuous variables not standardized

Stressors

- Our main stressor was IDACI, an LSOA level variable
- It aims to measure income deprivation affecting children
- Other stressors included:
  - FSM eligibility
  - House moves

Model

Variance functions for stress

Data

- Data is from previous cohort, who took GCSEs in 2006
- Postcodes with more than 2 students excluded
- Continuous variables not standardized

Stressors

- Our main stressor was IDACI, an LSOA level variable
- It aims to measure income deprivation affecting children
- Other stressors included:
  - FSM eligibility
  - House moves

Model

$$
\begin{aligned}
\mathbf{GCSE}_{ijk} = \alpha &+ \beta_1 \mathbf{pretest}_{ijk} + \beta_2 \mathbf{twin}_{jk} + \beta_5 \mathbf{stressor}_{jk} \\
&+ v_{0k} + u_{2jk}\mathbf{twin}_{jk} + e_{2ijk}\mathbf{twin}_{jk} + e_{3ijk}\mathbf{nontwin}_{jk} \\
&+ u_{4jk}\mathbf{twin} \cdot \mathbf{stressor}_{jk} + e_{4ijk}\mathbf{twin} \cdot \mathbf{stressor}_{jk} \\
&+ e_{6ijk}\mathbf{nontwin} \cdot \mathbf{stressor}_{jk}
\end{aligned}
$$

Variance functions for stress

Covariance structure

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{v0} \end{bmatrix}\right)$$

$$\begin{bmatrix} u_{2jk} \\ u_{4jk} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{u2} & \\ \sigma_{u24} & \sigma^2_{u4} \end{bmatrix}\right)$$

$$\begin{bmatrix} e_{2ijk} \\ e_{3ijk} \\ e_{4ijk} \\ e_{6ijk} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{e2} & & & \\ 0 & \sigma^2_{e3} & & \\ \sigma_{e24} & 0 & \sigma^2_{e4} & \\ 0 & \sigma_{e36} & 0 & \sigma^2_{e6} \end{bmatrix}\right)$$

Model

$$\begin{aligned} \textbf{GCSE}_{ijk} = \alpha &+ \beta_1 \textbf{pretest}_{ijk} + \beta_2 \textbf{twin}_{jk} + \beta_5 \textbf{stressor}_{jk} \\ &+ v_{0k} + u_{2jk}\textbf{twin}_{jk} + e_{2ijk}\textbf{twin}_{jk} + e_{3ijk}\textbf{nontwin}_{jk} \\ &+ u_{4jk}\textbf{twin} \cdot \textbf{stressor}_{jk} + e_{4ijk}\textbf{twin} \cdot \textbf{stressor}_{jk} \\ &+ e_{6ijk}\textbf{nontwin} \cdot \textbf{stressor}_{jk} \end{aligned}$$

Variance functions for stress

Covariance structure

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{v0} \end{bmatrix}\right)$$

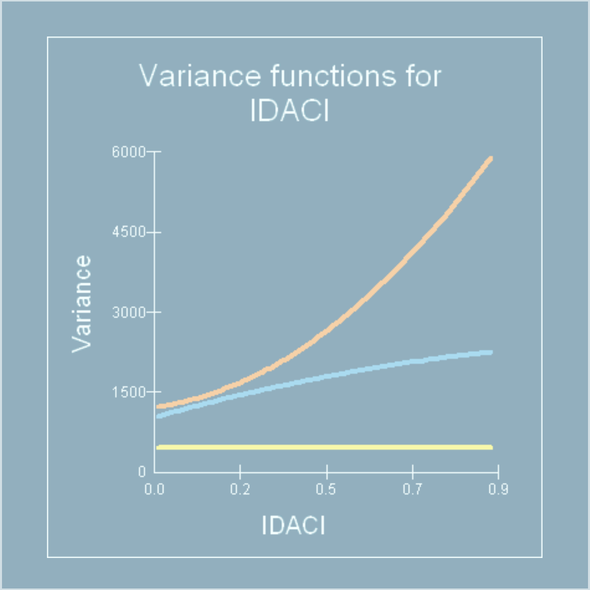
$$\begin{bmatrix} u_{2jk} \\ u_{4jk} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{u2} & \\ \sigma_{u24} & \sigma^2_{u4} \end{bmatrix}\right)$$

$$\begin{bmatrix} e_{2ijk} \\ e_{3ijk} \\ e_{4ijk} \\ e_{6ijk} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{e2} & & & \\ 0 & \sigma^2_{e3} & & \\ \sigma_{e24} & 0 & \sigma^2_{e4} & \\ 0 & \sigma_{e36} & 0 & \sigma^2_{e6} \end{bmatrix}\right)$$

Model

$$\begin{aligned} \textbf{GCSE}_{ijk} = \alpha &+ \beta_1 \textbf{pretest}_{ijk} + \beta_2 \textbf{twin}_{jk} + \beta_5 \textbf{IDACI}_{jk} \\ &+ v_{0k} + u_{2jk}\textbf{twin}_{jk} + e_{2ijk}\textbf{twin}_{jk} + e_{3ijk}\textbf{nontwin}_{jk} \\ &+ u_{4jk}\textbf{twin} \cdot \textbf{IDACI}_{jk} + e_{4ijk}\textbf{twin} \cdot \textbf{IDACI}_{jk} \\ &+ e_{6ijk}\textbf{nontwin} \cdot \textbf{IDACI}_{jk} \end{aligned}$$

Results

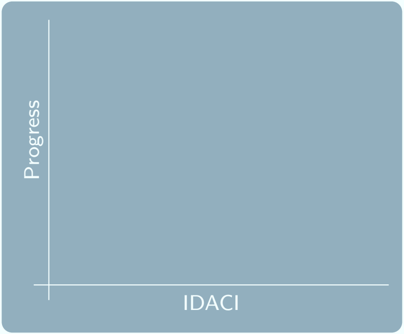


Variance functions for
IDACI
Variance
6000
4500
3000
1500
0
0.0
0.2
0.5
0.7
0.9
IDACI

Interpreting the results

As IDACI increases,

We have this situation:

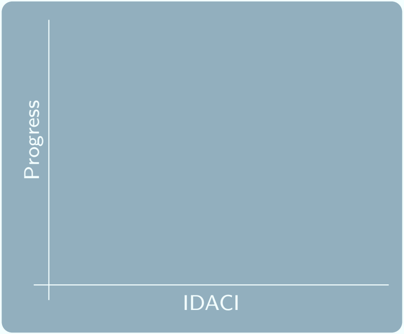


Interpreting the results

As IDACI increases,

- The mean progress decreases
  - $\beta_5 = -68.1$

We have this situation:

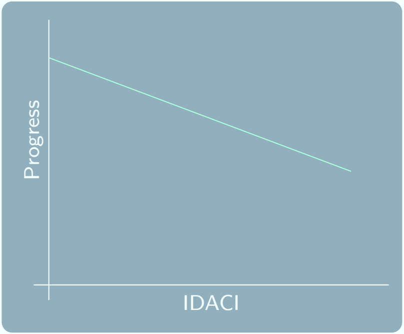


Interpreting the results

As IDACI increases,

- The mean progress decreases
  - $\beta_5 = -68.1$

We have this situation:

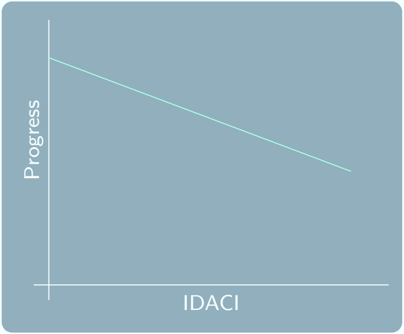


Interpreting the results

As IDACI increases,

- Between family variation increases
- The mean progress decreases
  - $\beta_5 = -68.1$

We have this situation:

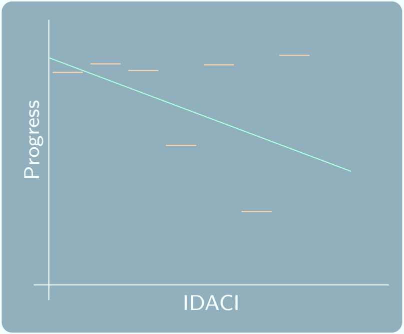


Interpreting the results

As IDACI increases,

- Between family variation increases
- The mean progress decreases
  - $\beta_5 = -68.1$

We have this situation:

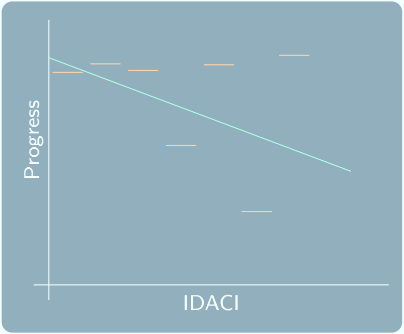


Interpreting the results

As IDACI increases,

- Between family variation increases
- Within family variation increases
- The mean progress decreases
  - $\beta_5 = -68.1$

We have this situation:

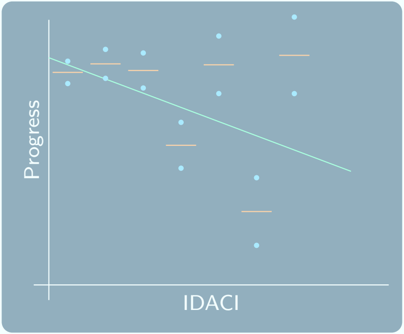


Interpreting the results

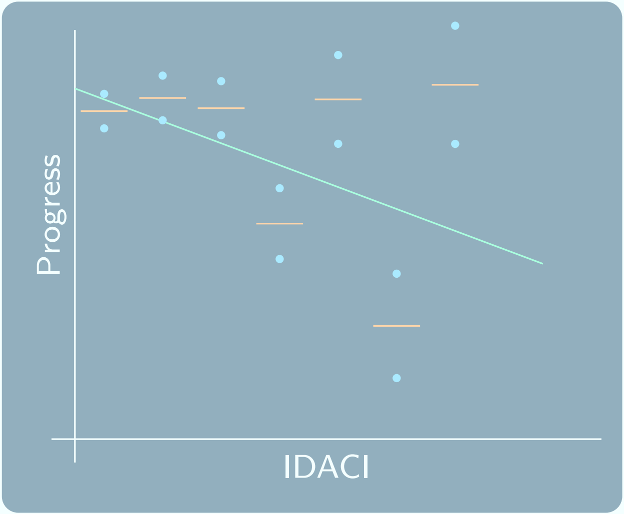
As IDACI increases,

- Between family variation increases
- Within family variation increases
- The mean progress decreases
  - $\beta_5 = -68.1$

We have this situation:



Interpreting the results

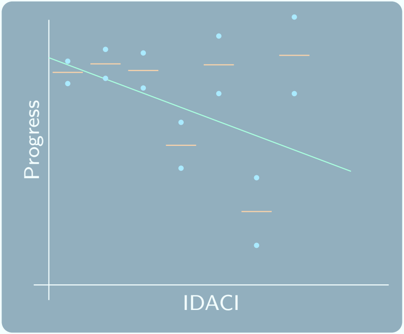


Interpreting the results

As IDACI increases,

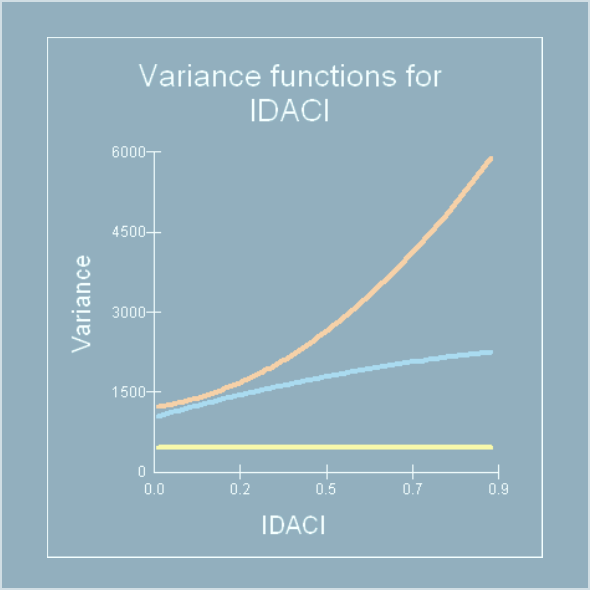
- Between family variation increases
- Within family variation increases
- The mean progress decreases
  - $\beta_5 = -68.1$

We have this situation:



- Between family variation increases more dramatically than within family variation
- So at greater levels of deprivation, family becomes relatively more important in determining progress

Results



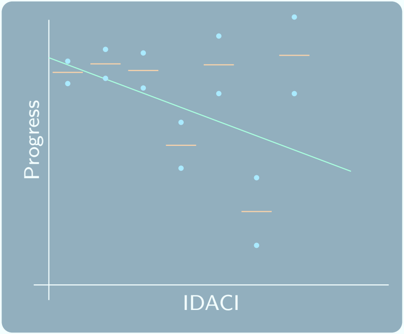
Variance functions for IDACI
Variance
6000
4500
3000
1500
0
0.0
0.2
0.5
0.7
0.9
IDACI

Interpreting the results

As IDACI increases,

- Between family variation increases
- Within family variation increases
- The mean progress decreases
  - $\beta_5 = -68.1$

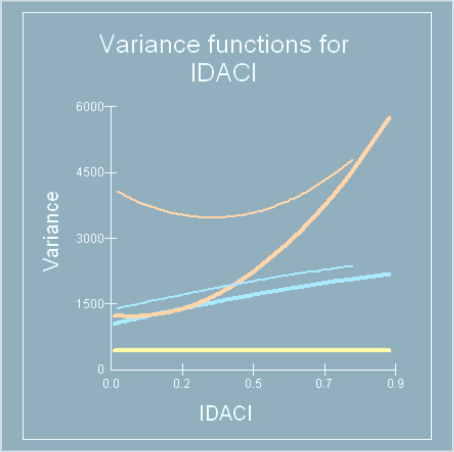
We have this situation:



- Between family variation increases more dramatically than within family variation
- So at greater levels of deprivation, family becomes relatively more important in determining progress

Other stressors

- We fitted the same model with different stressors:
  - IMD
  - FSM eligibility
  - Ever moved house
  - Number of house moves
  - Time since house move
- In almost all cases we see the same pattern
- We also fitted models with more than one stressor
  - e.g. IDACI and FSM eligibility
- In these models, both stressors show the same pattern



What's going on? Possible explanations

Genetic explanation

- Some families have genes which help to maintain progress in the presence of stressors, while others do not

Environmental explanation

What's going on? Possible explanations

Genetic explanation

- Some families have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability

Environmental explanation

What's going on? Possible explanations

Genetic explanation

- Some families have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some families have the gene and some don't

Environmental explanation

What's going on? Possible explanations

Genetic explanation

- Within families, some children have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some children have the gene and some don't

Environmental explanation

What's going on? Possible explanations

Genetic explanation

- Within families, some children have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some children have the gene and some don't

Environmental explanation

- Some families, across all levels of the stressors, have factors that make it harder to be good parents
  - alcoholism of parent
  - violent spouse

What's going on? Possible explanations

Genetic explanation

- Within families, some children have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some children have the gene and some don't

Environmental explanation

- Some families, across all levels of the stressors, have factors that make it harder to be good parents
  - alcoholism of parent
  - violent spouse
- In the absence of stressors, even families with these factors can provide a good environment for progress

What's going on? Possible explanations

Genetic explanation

- Within families, some children have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some children have the gene and some don't

Environmental explanation

- Some families, across all levels of the stressors, have factors that make it harder to be good parents
  - alcoholism of parent
  - violent spouse
- In the absence of stressors, even families with these factors can provide a good environment for progress
- In the presence of stressors, families with these factors cannot do so → variability since some families have these factors and some don't

What's going on? Possible explanations

Genetic explanation

- Within families, some children have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some children have the gene and some don't

Environmental explanation

- Children in families compete for resources

What's going on? Possible explanations

Genetic explanation

- Within families, some children have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some children have the gene and some don't

Environmental explanation

- Children in families compete for resources
- In the absence of stressors, there are enough resources for the needs of all children

What's going on? Possible explanations

Genetic explanation

- Within families, some children have genes which help to maintain progress in the presence of stressors, while others do not
- In the absence of a stressor, the genes make little difference so there is not much variability
- In the presence of a stressor, the genes make a big difference so there is variability arising from the fact that some children have the gene and some don't

Environmental explanation

- Children in families compete for resources
- In the absence of stressors, there are enough resources for the needs of all children
- In the presence of stressors, there are fewer resources and some children will have their needs met while others will not $\longrightarrow$ variability since those getting more resources can make more progress

References

References

- Bouchard, T. J. and MacGue, M. (1981) Familial Studies of Intelligence: A Review. *Science, New Series*, 212, 1055–1059
- Duncan, G. J., Boisjoly, J. and Mullan Harris, K. (2001) Sibling, peer, neighbourhood and schoolmate correlations as indicators of the importance of context for adolescent development. *Demography* **38(3)**: 437–447
- Fielding, A., Thomas, H., Steele, F., Browne, W., Leyland, A., Spencer, N. and Davison, I. (2006) Using Cross-Classified Multilevel Models to Improve Estimates of the Determination of Pupil Attainment: A Scoping Study. Research Report for Department for Education and Skills. School of Education, University of Birmingham, IBSN: 9780704426016 (International) 0704426013 (UK)
- Garner, C. and Raudenbush, S. W. (1991) Neighbourhood effects on educational attainment. *Sociology of Education*, **64**, 251–262
- Goldstein, H., Burgess, S. and McConnell, B. (2007) Modelling the effect of pupil mobility on school differences in educational achievement. *J. R. Statist. Soc. A*, **170**: 941–954
- Leckie, G. (2008) Modelling the effects of pupil mobility and neighbourhood on school differences in educational achievement. CMPO Working Paper 08/189
- Raudenbush, S. W. (1993) A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *J. Educ. Statist.*, **18**, 321–349
- Yang, M. and Woodhouse, G. (2001) Progress from GCSE to A and AS level: institutional and gender differences, and trends over time. *British Educational Research Journal*, **27**, 245–267