LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Familial and socio-economic influences on foetal growth across three generations

Bianca L De Stavola, David A Leon, Ilona Koupil*

London School of Hygiene and Tropical Medicine
and *University of Stockholm

LSHTM, 23 February 2011

Outline

Outline

Intergenerational correlations in size at birth

- Intergenerational studies: strong correlations in size at birth between parents and offspring **(0.20-0.25 for mothers, less for fathers)**.
- Two mechanisms:
  1. **Genetic**:
     - father-offspring: can only be explained via **foetal genes**
     - mother-offspring: could also be attributed to **maternal genes** that influence fetal growth
  2. **External environment**:
     - **maternal lifestyle** influences in utero environment (e.g. smoking and diet) and is correlated across generations.
- However:
  - **incorrect paternity** attribution may inflate difference between mother-offspring and father-offspring correlations
  - **limited evidence** of impact of earlier generations socio-economic factors on size at birth correlations
  - **little data** on more than 2 generations.

Intergenerational correlations in size at birth

- Intergenerational studies: strong correlations in size at birth between parents and offspring **(0.20-0.25 for mothers, less for fathers)**.
- Two mechanisms:
  1. **Genetic**:
     - father-offspring: can only be explained via **foetal genes**
     - mother-offspring: could also be attributed to **maternal genes** that influence fetal growth
  2. **External environment**:
     - **maternal lifestyle** influences in utero environment (e.g. smoking and diet) and is correlated across generations.
- However:
  - **incorrect paternity** attribution may inflate difference between mother-offspring and father-offspring correlations
  - **limited evidence** of impact of earlier generations socio-economic factors on size at birth correlations
  - **little data** on more than 2 generations.

Intergenerational correlations in size at birth

- Intergenerational studies: strong correlations in size at birth between parents and offspring **(0.20-0.25 for mothers, less for fathers)**.
- Two mechanisms:
  1. **Genetic**:
     - father-offspring: can only be explained via **foetal genes**
     - mother-offspring: could also be attributed to **maternal genes** that influence fetal growth
  2. **External environment**:
     - **maternal lifestyle** influences in utero environment (e.g. smoking and diet) and is correlated across generations.
- However:
  - **incorrect paternity** attribution may inflate difference between mother-offspring and father-offspring correlations
  - **limited evidence** of impact of earlier generations socio-economic factors on size at birth correlations
  - **little data** on more than 2 generations.

Intergenerational correlations in size at birth

- Intergenerational studies: strong correlations in size at birth between parents and offspring **(0.20-0.25 for mothers, less for fathers)**.
- Two mechanisms:
  1. **Genetic**:
     - father-offspring: can only be explained via **foetal genes**
     - mother-offspring: could also be attributed to **maternal genes** that influence fetal growth
  2. **External environment**:
     - **maternal lifestyle** influences in utero environment (e.g. smoking and diet) and is correlated across generations.
- However:
  - **incorrect paternity** attribution may inflate difference between mother-offspring and father-offspring correlations
  - **limited evidence** of impact of earlier generations socio-economic factors on size at birth correlations
  - **little data** on more than 2 generations.

Aims

The **aims** of the talk are to assess:

A) whether correlations in size at birth across three generations are consistent with those found across two generations

B) the extent to which these correlations could be explained by socio-demographic continuities across generations.

Using the unique and rich data available in UBCoS Multigen

Uppsala Birth Cohort Multigenerational Study (UBCoS Multigen)

Prospective study of men and women born in Uppsala, Sweden (1915-1929) and their descendants (currently linking to **4th & 5th** generation)

Am J Epidemiol, to appear

Aims

The **aims** of the talk are to assess:

A) whether correlations in size at birth across three generations are consistent with those found across two generations

B) the extent to which these correlations could be explained by socio-demographic continuities across generations.

Using the unique and rich data available in UBCoS Multigen

Uppsala Birth Cohort Multigenerational Study (UBCoS Multigen)

Prospective study of men and women born in Uppsala, Sweden (1915-1929) and their descendants (currently linking to **4th & 5th** generation)

Am J Epidemiol, to appear

Aims

The **aims** of the talk are to assess:

A) whether correlations in size at birth across three generations are consistent with those found across two generations

B) the extent to which these correlations could be explained by socio-demographic continuities across generations.

Using the unique and rich data available in UBCoS Multigen

Uppsala Birth Cohort Multigenerational Study (UBCoS Multigen)

Prospective study of men and women born in Uppsala, Sweden (1915-1929) and their descendants (currently linking to **4th & 5th** generation)

Am J Epidemiol, to appear

Outline

Three generations

- **Original cohort** (Leon et al. 1998)
  - Generation 1 (G1): all Uppsala births in 1915-29 ($\backsim$ 14,000),
    Source: Uppsala Academic Hospital (UAH) birth records
- **descendants traced via the Swedish Multigenerational Registry** (Koupil, 2007)**:**
  - Generation 2 (G2): their children ($\backsim$ 20,000),
  - Generation 3 (G3): their grandchildren ($\backsim$ 33,000)
  - great-grand-children (G4), great-great-grand-children (G5) ...
    Source: several
- **8,550 UBCoS G1 grandparents with 33,693 grandchildren**

Three generations

- **Original cohort** (Leon et al. 1998)
  - Generation 1 (G1): all Uppsala births in 1915-29 ($\backsim$ 14,000),
    Source: Uppsala Academic Hospital (UAH) birth records
- **descendants traced via the Swedish Multigenerational Registry** (Koupil, 2007)**:**
  - Generation 2 (G2): their children ($\backsim$ 20,000),
  - Generation 3 (G3): their grandchildren ($\backsim$ 33,000)
  - great-grand-children (G4), great-great-grand-children (G5) ...
    Source: several
- **8,550 UBCoS G1 grandparents with 33,693 grandchildren**

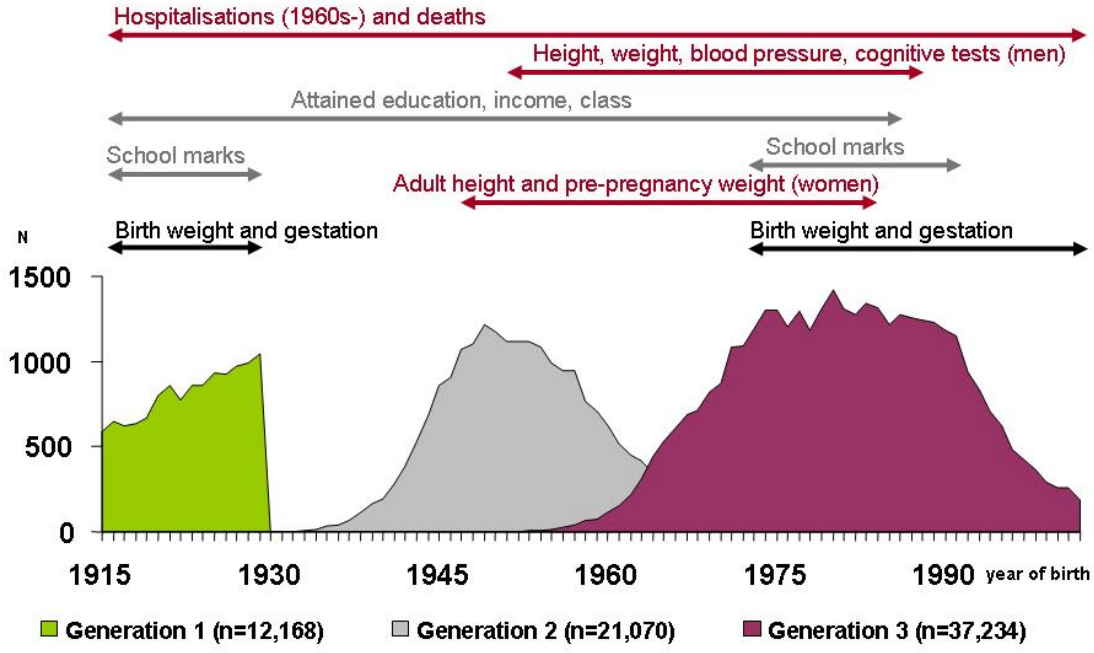
Three generations

- **Original cohort** (Leon et al. 1998)
  - Generation 1 (G1): all Uppsala births in 1915-29 ($\backsim$ 14,000),
    Source: Uppsala Academic Hospital (UAH) birth records
- **descendants traced via the Swedish Multigenerational Registry** (Koupil, 2007)**:**
  - Generation 2 (G2): their children ($\backsim$ 20,000),
  - Generation 3 (G3): their grandchildren ($\backsim$ 33,000)
  - great-grand-children (G4), great-great-grand-children (G5) ...
    Source: several
- **8,550 UBCoS G1 grandparents with 33,693 grandchildren**

Data sources

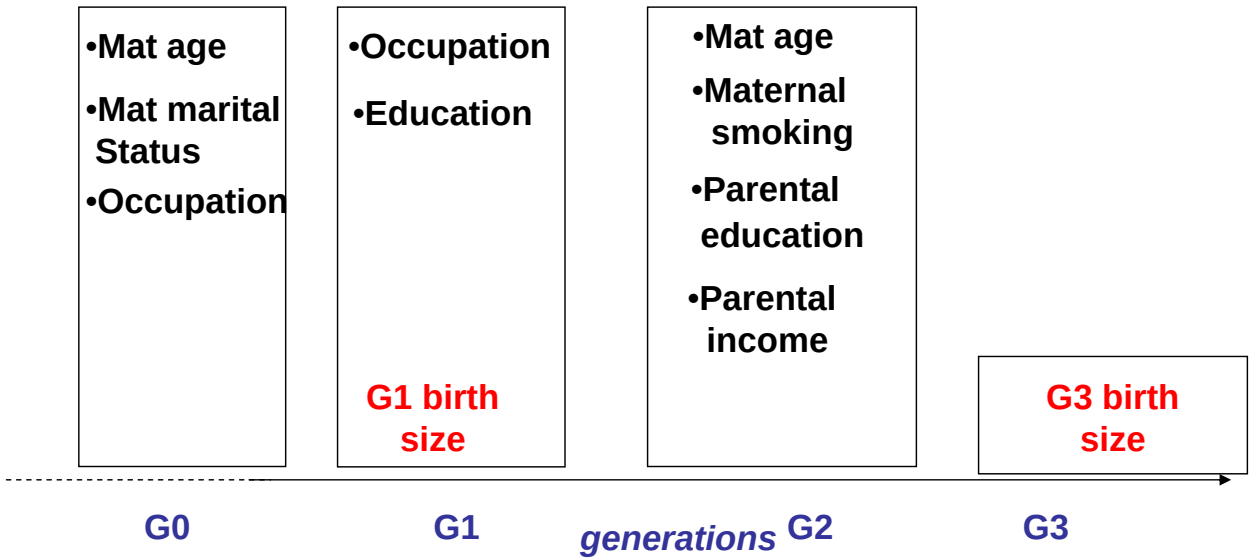
- **Birth data:** UAH, Swedish Birth Registry (1972-)
- **SEP, demographic vars** : Censuses (1960, 1970, 1980), linked longitudinal studies, etc
- **Others**: Conscripts Register, School records, ...

Linkage depends on calendar year $\Rightarrow$ no birth data for G2



Main variables and data for these analyses

- **standardized size at birth** (birth weight and length)
- **SEP and demographic variables**



- **inclusions**, for each generation: singletons, not adoptees
- **restrictions**: to G3 with birth data $\Rightarrow$ **7,657** G1 and **25,141** G3

Data complexities

Four types of grandparents:

GP type	G1	G1-G3
Mat GMoth	2,340	6,169
Mat GFath	2,612	7,219
Pat GMoth	2,490	6,965
Mat GFath	2,694	7,799
Total	10,136	**28,152**
All Gpar	$7,657^{*}$	$25,141^{\dagger}$

* Some grandparents contributes to maternal and paternal entries

† Some grandchildren have more than one grandparent

Analytical complexity:

(a) **Clustered data**

- $\sim$ 3 G3 per G1

(b) **Prospective design**

- Incomplete pedigrees

(c) **Missing data**

- Mat Gmothers least complete

Data complexities

Four types of grandparents:

GP type	G1	G1-G3
Mat GMoth	2,340	6,169
Mat GFath	2,612	7,219
Pat GMoth	2,490	6,965
Mat GFath	2,694	7,799
Total	10,136	**28,152**
All Gpar	$7,657^{*}$	$25,141^{\dagger}$

* Some grandparents contributes to maternal and paternal entries

† Some grandchildren have more than one grandparent

Analytical complexity:

(a) **Clustered data**

- $\sim$ 3 G3 per G1

(b) **Prospective design**

- Incomplete pedigrees

(c) **Missing data**

- Mat Gmothers least complete

Data complexities

Four types of grandparents:

GP type	G1	G1-G3
Mat GMoth	2,340	6,169
Mat GFath	2,612	7,219
Pat GMoth	2,490	6,965
Mat GFath	2,694	7,799
Total	10,136	**28,152**
All Gpar	$7,657^{*}$	$25,141^{\dagger}$

* Some grandparents contributes to maternal and paternal entries

† Some grandchildren have more than one grandparent

Analytical complexity:

(a) **Clustered data**

- $\sim$ 3 G3 per G1

(b) **Prospective design**

- Incomplete pedigrees

(c) **Missing data**

- Mat Gmothers least complete

Data complexities

Four types of grandparents:

GP type	G1	G1-G3
Mat GMoth	2,340	6,169
Mat GFath	2,612	7,219
Pat GMoth	2,490	6,965
Mat GFath	2,694	7,799
Total	10,136	**28,152**
All Gpar	$7,657^{*}$	$25,141^{\dagger}$

* Some grandparents contributes to maternal and paternal entries

† Some grandchildren have more than one grandparent

Analytical complexity:

(a) **Clustered data**

- $\sim$ 3 G3 per G1

(b) **Prospective design**

- Incomplete pedigrees

(c) **Missing data**

- Mat Gmothers least complete

Outline

A) are correlations maintained across 3 generations?

Correlation coefficients for standardized BW (N_{pairs}=28,152)

	4 groups		2 groups	
G1	**Coef.**	**(95% CI)**	**Coef.**	**(95% CI)**
Mat gm	0.124	(0.095, 0.153)	0.125	(0.105, 0.145)
Mat gf	0.126	(0.099, 0.153)		
Pat gm	0.093	(0.065, 0.121)	0.096	(0.077, 0.115)
Pat gf	0.099	(0.073, 0.126)		

Estimates based on a RE model with GP as clusters; adj for G1 & G3 parity and year of birth

- **Correlations maintained across generations**
- Stronger associations for maternal grandparents
- Model with 2 groups equally good fit → maternal/paternal lineage as the main discriminant (p=*0.02*)

B) are correlations 'explained' by socio-dem continuities?

Including SEP variables as potential **mediators**:

	Minimally adjusted		Fully adjusted	
	Cor. Coef.	**(95% CI)**	**Cor. Coef.**	**(95% CI)**
Mat GP	0.122	(0.095, 0.150)	0.121	(0.093, 0.148)
Pat GP	0.091	(0.065, 0.117)	0.092	(0.066, 0.117)

GP: grandparent

- Minimally: adjusted (for G0 & G2 parity and G1 & G3 year of birth)
- Fully: additionally adjusted for G2 smoking, G2 income, G1 and G2 education, G0 and G2 mat age , G0 and G1 SEP, G0 mat marital status
- analysis carried out on a subset because of missing values: N_{pairs}=14,382

No evidence of mediation (or effect modification)

Critique

Results might be affected by bias because:

- likely measurement/missclassification error affecting the socio-economic variables
- missing data

They also do not exploit the family structure of the data.

Alternative approach:

Genetic biometrical model

Partition total variance of size at birth into: foetal genes, maternal genes, shared environmental factors, unshared environmental factors.

Critique

Results might be affected by bias because:

- likely measurement/missclassification error affecting the socio-economic variables
- missing data

They also do not exploit the family structure of the data.

Alternative approach:

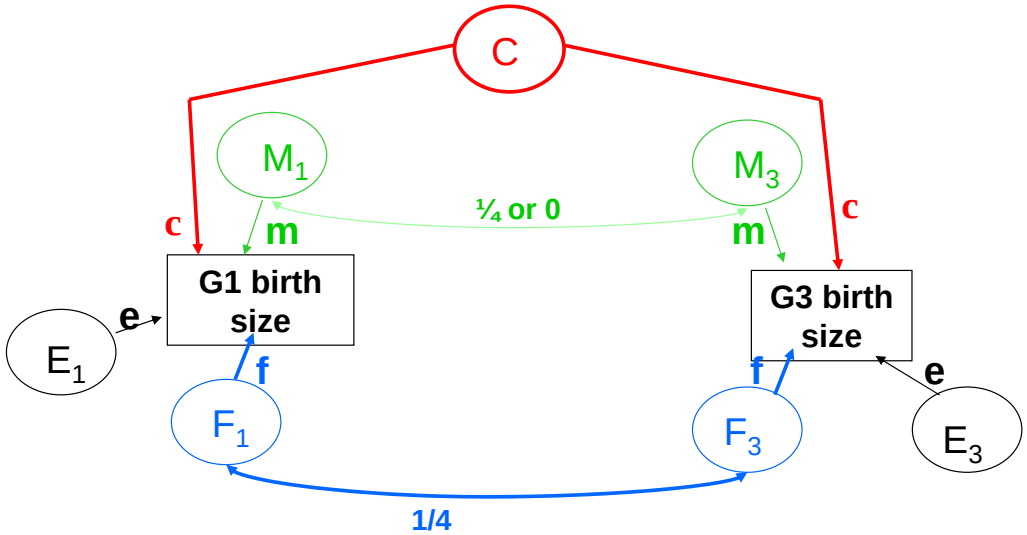
Genetic biometrical model

Partition total variance of size at birth into: foetal genes, maternal genes, shared environmental factors, unshared environmental factors.

Outline

A model for the UBCoS data

Partitioning the variance of size at birth of the G1 and G3 study members, under certain assumptions, may lead to a model such as this:



Ass: rand mating, genotypic parent-child: 0.5; no interactions, constant effects

An introduction using twins data

Twin design:

- Data on **Y**: phenotype of interest
- measured in pairs of monozygotic (**MZ**) and dizygotic (**DZ**) twins
- Exploit: **MZ** twins share all their genes while **DZ** share half

ACE Model

Widely used to separate genetic from environmental sources of variation in a phenotype **Y**.

Specific assumptions:

- genes have additive effects
- twins in a set experience the same environment (at least in childhood)

Gielen et al, Behav Genet (2008) 38:4454

An introduction using twins data

Twin design:

- Data on **Y**: phenotype of interest
- measured in pairs of monozygotic (**MZ**) and dizygotic (**DZ**) twins
- Exploit: **MZ** twins share all their genes while **DZ** share half

ACE Model

Widely used to separate genetic from environmental sources of variation in a phenotype **Y**.

Specific assumptions:

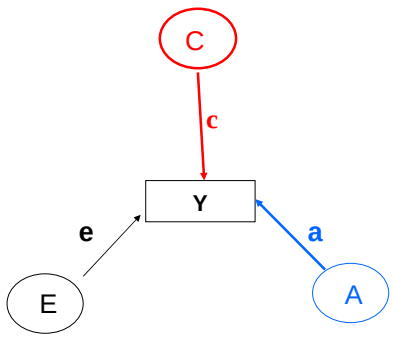
- genes have additive effects
- twins in a set experience the same environment (at least in childhood)

Gielen et al, Behav Genet (2008) 38:4454

The ACE model for one member of the twins set

The phenotype is a weighted sum of three components:

$$\mathbf{Y = aA + cC + eE}$$

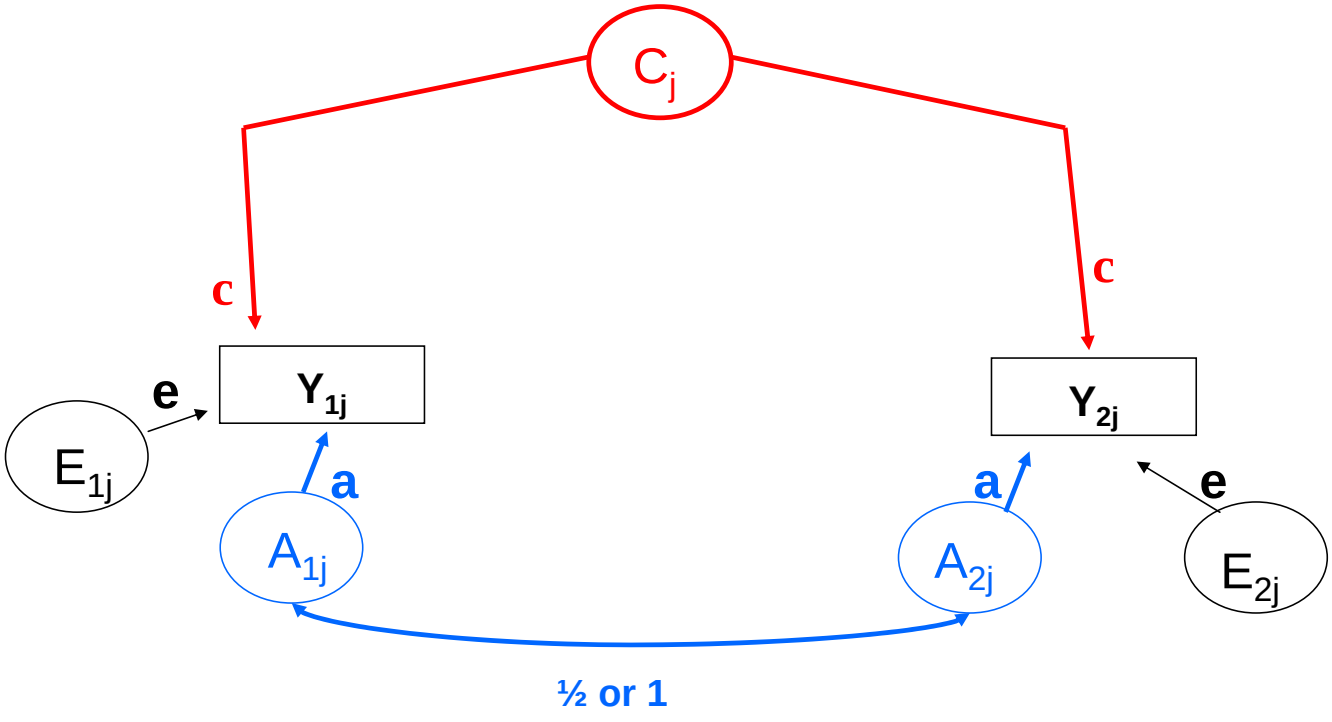


- **genetic** factors: represented by the latent variable *A*
- **shared** environmental factors: represented by **C**
- **unshared** environmental factors: represented by **E**

All latent factors: independent N(0,1); a, c, and e: *path coefficients*.

The ACE model for the twins set

$\mathbf{Y}_{ij}$: phenotype for twin i in set j



Covariance between $\mathbf{A_{1j}}$ and $\mathbf{A_{2j}}$ is 1 for **MZ** and $\frac{1}{2}$ for **DZ**

Heritability and identification

$\mathbf{Y_{ij} = aA_{ij} + cC_j + eE_{ij}}$
This model implies:

- $Var(Y_{ij}) = a^2 + c^2 + e^2$
- $Cov(Y_{1j}, Y_{2j}) = a^2 + c^2$ for **MZ** twins
- $Cov(Y_{1j}, Y_{2j}) = \frac{a^2}{2} + c^2$ for **DZ** twins

Useful:

to estimate heritability (h)$=\frac{a^2}{a^2+c^2+e^2}$

Identification:

The model has 3 parameters with three sufficient statistics

Heritability and identification

$\mathbf{Y_{ij} = aA_{ij} + cC_j + eE_{ij}}$
This model implies:

- $Var(Y_{ij}) = a^2 + c^2 + e^2$
- $Cov(Y_{1j}, Y_{2j}) = a^2 + c^2$ for **MZ** twins
- $Cov(Y_{1j}, Y_{2j}) = \frac{a^2}{2} + c^2$ for **DZ** twins

Useful:

to estimate heritability (h)$=\frac{a^2}{a^2+c^2+e^2}$

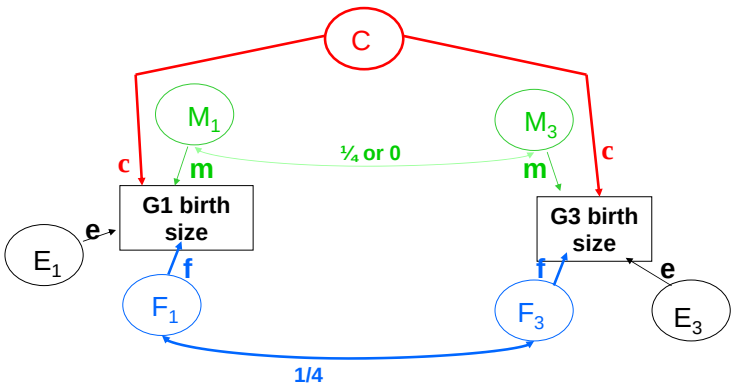
Identification:

The model has 3 parameters with three sufficient statistics

An extension for G1-G3 birth size data

Partition the variance of standardised BW into components from:

- **foetal** genes: represented by F; **maternal** genes: M
- **shared** environm factors: C; **unshared** environmental factors: E



Ass: rand mating, genotypic parent-child: 0.5; no interactions, constant effects

Covariance btwn F_{1j} and F_{3j} is $\frac{1}{4}$ for all, covariance btwn M_{1j} and M_{3j} is $\frac{1}{4}$ for Mat Grandparents, 0 otherwise.

Genetic biometrical model

Identification

$\mathbf{Y}_{ij}$: phenotype for member i in family j

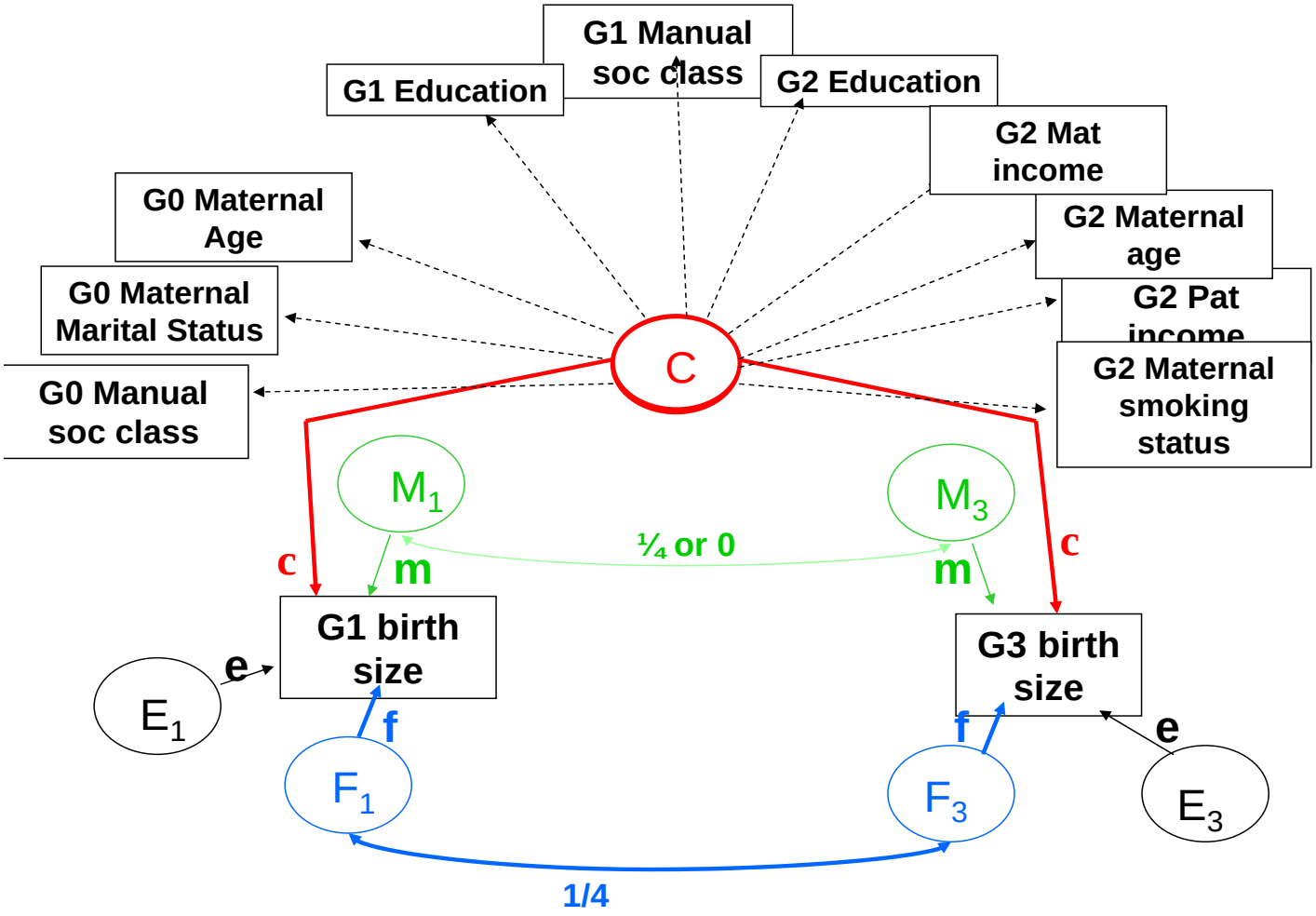
$$\mathbf{Y_{ij}} = \mathbf{fF_{ij}} + \mathbf{mM_{ij}} + \mathbf{cC_j} + \mathbf{eE_{ij}}$$

This model implies:

- $Var(Y_{ij}) = f^2 + m^2 + c^2 + e^2$
- $Cov(Y_{1j}, Y_{3j}) = \frac{1}{4}f^2 + \frac{1}{4}m^2 + c^2$ for Maternal GParents
- $Cov(Y_{1j}, Y_{3j}) = \frac{1}{4}f^2 + c^2$ for Paternal GParents

Since the model has 4 parameters, it is not identified.
However, it can be identified if we specify **C** in terms of some observables

The full genetic biometrical model for G1-G3 pairs



Genetic biometrical model

Contributions to the variances and covariances

The model implies:

- $Var(Y_{ij}) = f^2 + m^2 + c^2 + e^2$
- $Cov(Y_{1j}, Y_{3j}) = \frac{1}{4}f^2 + \frac{1}{4}m^2 + c^2$ for Maternal GParents
- $Cov(Y_{1j}, Y_{3j}) = \frac{1}{4}f^2 + c^2$ for Paternal GParents

Contributions to the variance:

- *Foetal* $= \frac{f^2}{f^2+m^2+c^2+e^2}$
- *Maternal* $= \frac{m^2}{f^2+m^2+c^2+e^2}$
- *Shared environment* $= \frac{c^2}{f^2+m^2+c^2+e^2}$

Contribution to the covariances:

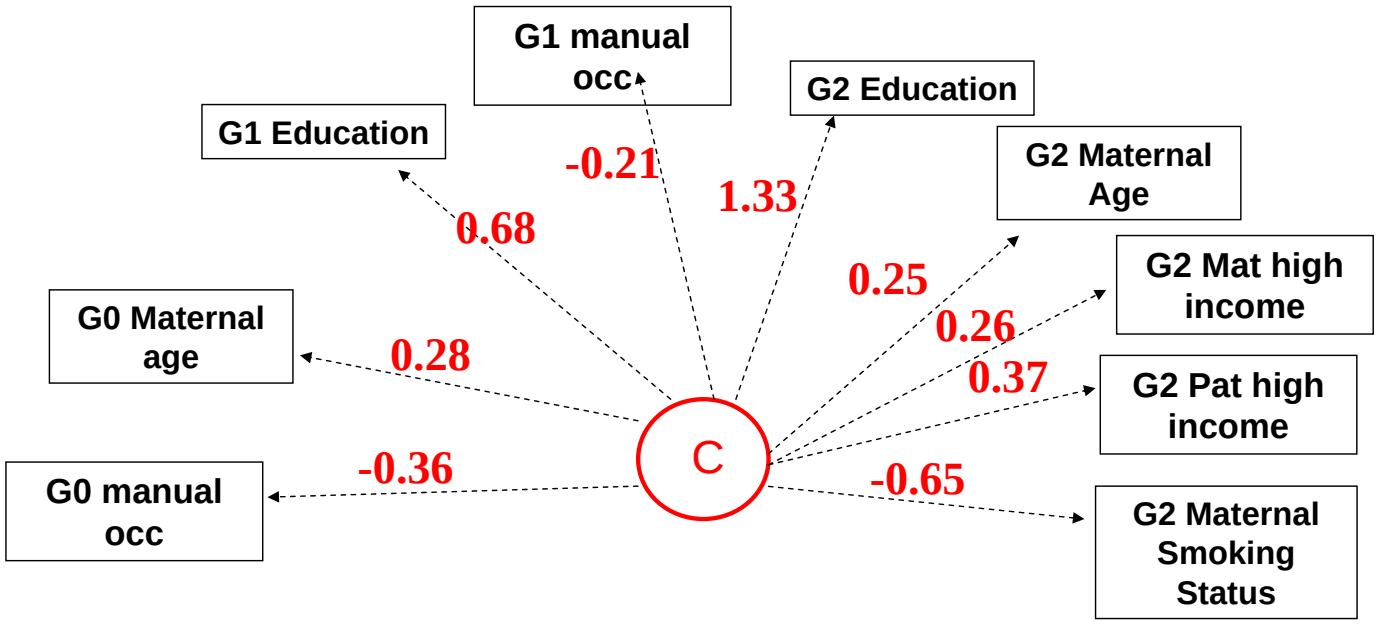
$\Rightarrow$ *Environm contribution of Maternal GP*: $\frac{c^2}{f^2/4+m^2/4+c^2}$

$\Rightarrow$ *Environm contribution of Paternal GP*: $\frac{c^2}{f^2/4+c^2}$

When Y is standardized, these are contributions to the correlations

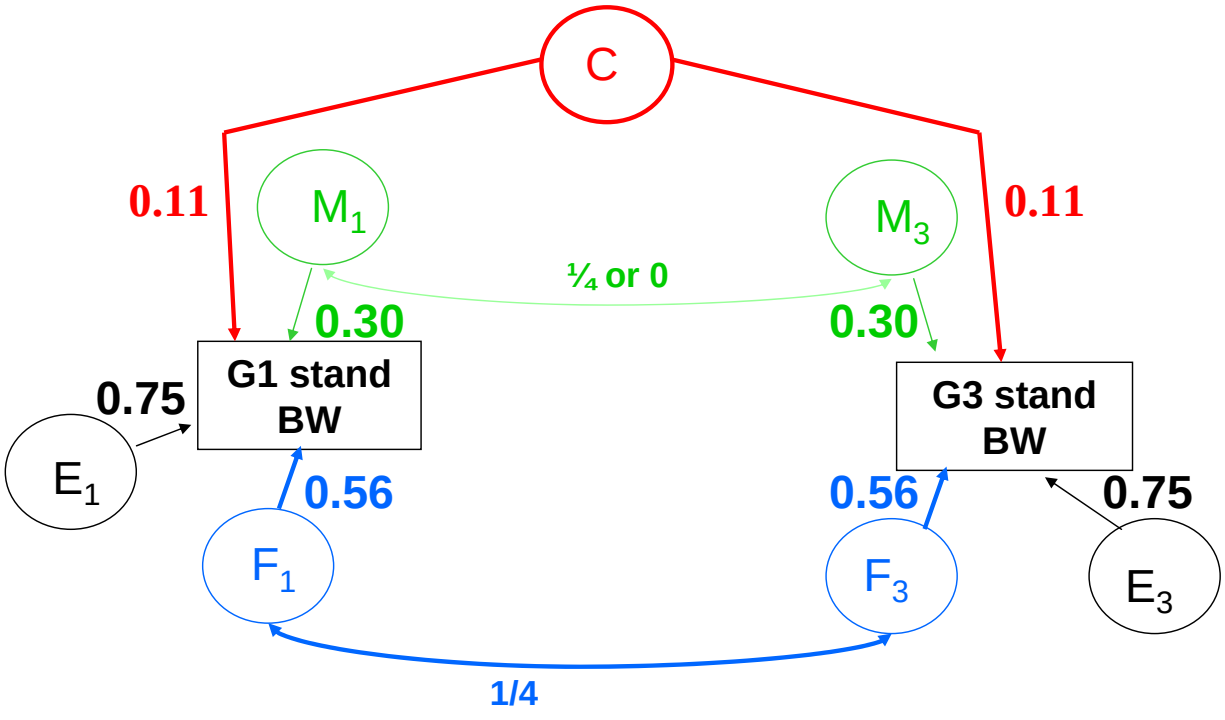
Outline

Factor loadings for C



Higher socio-demographic status ⇒ positive C

Path coefficients



- **Good fit**: Observed \ predicted correlations: **Mat GP**: 0.114 \ 0.113; **Pat GP**: 0.093 \ 0.090
- Estimates conditional on G0 and G2 parity and G1 and G3 year of birth
- robust SE to account for clustering: $N_{min} = 14\,382$

Contributions to the correlations

	Stand Birth weight		Stand Birth length	
	Estimate	95% CI	Estimate	95% CI
Mat Environment	11.6	5.9, 17.3	13.0	6.1, 20.0
Pat Environment	13.9	7.0, 20.8	14.6	7.0, 22.1

- Estimates of fetal, maternal and shared environment contributions to the variance: $\sim$ 30% ; $\sim$10%; $\sim$1%.

Outline

Linkage leads to systematic missingness

	Variable	Missing
G2	Parental education, income	1-2%
	Maternal age	0%
	Parity	0%
	Maternal smoking	46%
G1	Education, SEP	0-2%
	Parity	3%
G0	Mat marital status	0%
	Maternal age	0%
All		49%

Mat grandmother most affected by missingness

Sources of missingness

- Due to Linkage:
  missing maternal smoking status because born before 1980 (when smoking status started to be recorded)

  $\Rightarrow$ G3 Year of birth 'causes' missingness
- G3 year of birth is completely observed: $\Rightarrow$ MAR
- G3 year of birth is already included in the model: $\Rightarrow$ Missing mechanism is ignorable if using ML+EM

Sources of missingness

- Due to Linkage:
  missing maternal smoking status because born before 1980 (when smoking status started to be recorded)

  $\Rightarrow$ G3 Year of birth 'causes' missingness
- G3 year of birth is completely observed: $\Rightarrow$ MAR
- G3 year of birth is already included in the model: $\Rightarrow$ Missing mechanism is ignorable if using ML+EM

Sources of missingness

- Due to Linkage:
  missing maternal smoking status because born before 1980 (when smoking status started to be recorded)

  $\Rightarrow$ G3 Year of birth 'causes' missingness
- G3 year of birth is completely observed: $\Rightarrow$ MAR
- G3 year of birth is already included in the model: $\Rightarrow$ Missing mechanism is ignorable if using ML+EM

Updated contributions to the correlations

	Stand Birth weight		Stand Birth length	
	Estimate	95% CI	Estimate	95% CI
ML under MCAR				
Mat Environment	11.6	5.9, 17.3	13.0	6.1, 20.0
Pat Environment	13.9	7.0, 20.8	14.6	7.0, 22.1
ML+EM under MAR				
Mat Environment	15.8	10.7, 21.0	19.2	12.6, 25.8
Pat Environment	20.6	13.6, 27.7	19.5	13.1, 25.8

Evidence of contribution of shared environment

Revisiting the regression models

Replacing the multiple environmental indicators with estimated **C**:

	Minim adjusted		Fully adjusted		C adjusted	
	Cor.	**(95% CI)**	**Cor.**	**(95% CI)**	**Cor.**	**(95% CI)**
Mat GP	0.122	(0.095, 0.150)	0.121	(0.093, 0.148)	0.116	(0.089, 0.144)
Pat GP	0.091	(0.065, 0.117)	0.092	(0.066, 0.117)	0.087	(0.061, 0.113)

GP: grandparent

Some effect is now explained by the environmental factors.

Outline

Specific comments

Do continuity in size at birth depend on social disadvantage?

- shared environment 13-19% of the intergenerational correlations in standardized size at birth

On which basis?

- Assuming a specific genetic model, in particular a specific maternal genetic influence
- Assuming missingness was MAR, estimates were slightly inflated
- Results robust when some of the assumptions were relaxed

Why are results different from those from standard regression?

- addressing measurement error in SEP/demographic indicators recuperated part of the effect of shared environment not identified by standard regression

General comments

- Separating biological and social pathways involves assuming and measuring specific pathways
- When these cannot be properly measured an alternative approach is to separate components of 'correlations' into biological and environmental pathways
- Explicitly assessing the interplay of biology and environment would require access to genetic data (and several other assumptions)
- Final results are very much dependent on the quality of the data

Selected References

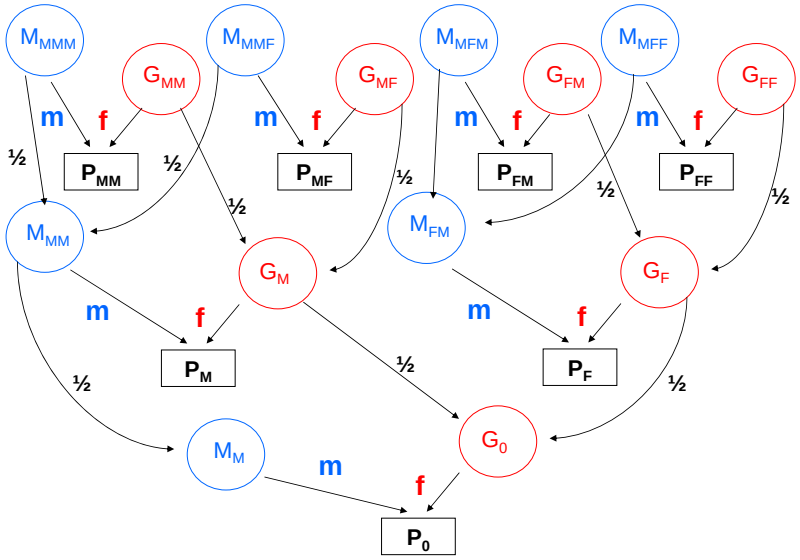
- Bollen. *Structural equations with latent variables*. 1989. New York: John Wiley.
- Goodman A, Koupil I. Social and biological determinants of reproductive success in Swedish males and females born 1915-1929. *Evolution & Human Behavior* 2009;30:329-41.
- Leon DA et al. Reduced fetal growth rate and increased risk of death from ischaemic heart disease: cohort study of 15 000 Swedish men and women born 1915-29. *Br Med J* 1998; 317:241-5.
- Lunde et al. Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent offspring data. *AJE* 2007; 165: 734-741.
- Magnus et al. Paternal contribution to birth weight. *J Epidemiol Community Health* 2001; 55: 873-877.
- Skrondal A, Rabe-Hesketh S. *Generalised latent variable modeling*. 2004. London: Chapman & Hall.

Variance components

	Stand Birth weight		Stand Birth length	
	Estimate	95% CI	Estimate	95% CI
Foetal genetic	32.5	21.4,43.5	26.6	16.6, 36.6
Maternal genetic	7.5	-9.6,24.5	3.7	-11.9,19.2
Environment	1.3	0.8, 1.9	1.1	0.1, 1.6

- Weaker results for head circumference
- Similar results for birth weight and birth length, adjusted for gest age and sex
- Reassuring that estimates of genetic and maternal heritability similar to those published by Lunde (2007)

The 3 generations genetic model



P: phenotype

M: maternal gene

G: foetal gene

Potential biases

Misclassification

What if paternity is erroneously attributed?